Lecture Notes in Electrical Engineering 421

James J. (Jong Hyuk) Park
Yi Pan
Gangman Yi
Vincenzo Loia
Editors

Advances in Computer Science and Ubiquitous Computing

CSA-CUTE2016

Springer

Lecture Notes in Electrical Engineering

Volume 421

Board of Series editors

Leopoldo Angrisani, Napoli, Italy
Marco Arteaga, Coyoacán, México
Samarjit Chakraborty, München, Germany
Jiming Chen, Hangzhou, P.R. China
Tan Kay Chen, Singapore, Singapore
Rüdiger Dillmann, Karlsruhe, Germany
Haibin Duan, Beijing, China
Gianluigi Ferrari, Parma, Italy
Manuel Ferre, Madrid, Spain
Sandra Hirche, München, Germany
Faryar Jabbari, Irvine, USA
Janusz Kacprzyk, Warsaw, Poland
Alaa Khamis, New Cairo City, Egypt
Torsten Kroeger, Stanford, USA
Tan Cher Ming, Singapore, Singapore
Wolfgang Minker, Ulm, Germany
Pradeep Misra, Dayton, USA
Sebastian Möller, Berlin, Germany
Subhas Mukhopadyay, Palmerston, New Zealand
Cun-Zheng Ning, Tempe, USA
Toyoaki Nishida, Sakyo-ku, Japan
Bijaya Ketan Panigrahi, New Delhi, India
Federica Pascucci, Roma, Italy
Tariq Samad, Minneapolis, USA
Gan Woon Seng, Nanyang Avenue, Singapore
Germano Veiga, Porto, Portugal
Haitao Wu, Beijing, China
Junjie James Zhang, Charlotte, USA

About this Series

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

James J. (Jong Hyuk) Park
Yi Pan · Gangman Yi · Vincenzo Loia
Editors

Advances in Computer Science and Ubiquitous Computing

CSA-CUTE2016

Springer

Editors
James J. (Jong Hyuk) Park
Computer Science and Engineering
Seoul National University of Science and Technology
Seoul
Korea (Republic of)

Yi Pan
Department of Computer Science
Georgia State University
Atlanta, GA
USA

Gangman Yi
Computer Science and Engineering
Gangneung-Wonju National University
Wonju
Korea (Republic of)

Vincenzo Loia
University Salerno
Fisciano
Italy

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-3022-2 ISBN 978-981-10-3023-9 (eBook)
DOI 10.1007/978-981-10-3023-9

Library of Congress Control Number: 2015957230

© Springer Nature Singapore Pte Ltd. 2017
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.
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.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #22-06/08 Gateway East, Singapore 189721, Singapore

Message from the CSA 2016 General Chair

International Conference on Computer Science and its Applications (CSA 2016) is the eighth event in the series of international scientific conference. This conference takes place in Bangkok, Thailand, during December 19–21, 2016. CSA 2016 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications. CSA 2016 will provide an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of CSA. In addition, the conference will publish high-quality papers which are closely related to the various theories and practical applications in CSA. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA 2016 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA 2015 (7th Edition: Cebu, Philippines, December 2015), CSA 2014 (6th Edition: Guam, December, 2014), CSA 2013 (5th Edition: Danang, December, 2013), CSA 2012 (4th Edition: Jeju, November, 2012), 1CSA 2011 (3rd Edition: Jeju, December, 2011), CSA 2009 (2nd Edition: Jeju, December, 2009), and CSA 2008 (1st Edition: Australia, October, 2008).

The papers included in the proceedings cover the following topics: Mobile and ubiquitous computing, Dependable, reliable and autonomic computing, Security and trust management, Multimedia systems and services, Networking and communications, Database and data mining, Game and software engineering, Grid and scalable computing, Embedded system and software, Artificial intelligence, Distributed and parallel algorithms, Web and Internet computing and IT policy and business management.

Accepted and presented papers highlight new trends and challenges in the area of computer science and its applications. The presenters showed how new research could lead to novel and innovative applications. We hope you will find these results useful and inspiring for your future research. We would like to express our sincere thanks to Steering Chairs: Prof. James J. (Jong Hyuk) Park (SeoulTech, Korea), Prof. Han-Chieh Chao (National Ilan University, Taiwan) and Prof. Mohammad S. Obaidat (Monmouth University, USA). Our special thanks go to the Program

Chairs: Prof. Gangman Yi (Dongguk University, Korea), Prof. Houcine Hassan (Universitat Politècnica de València, Spain), Prof. Yu Chen (State University of New York, USA), Prof. Hwamin Lee (SoonChunHyang University, Korea), Prof. Jin Wang (Yangzhou University, China), all Program Committee members and all the additional reviewers for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

We cordially thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

CSA 2016 General Chair
Young-Sik Jeong
Yang Xiao
Vincenzo Loia
Rung-Shiang Cheng
Victor Leung
Ka Lok Man

Message from the CSA 2016 Program Chairs

Welcome to the 8th International Conference on Computer Science and its Applications (CSA 2016) which will be held in Bangkok, Thailand, during December 19–21, 2016. CSA 2016 will be the most comprehensive conference focused on the various aspects of advances in computer science and its applications.

CSA 2016 provides an opportunity for academic and industry professionals to discuss the latest issues and progress in the area of computer science. In addition, the conference contains high-quality papers which are closely related to the various theories and practical applications in computer science. Furthermore, we expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject. CSA 2016 is the next event in a series of highly successful International Conference on Computer Science and its Applications, previously held as CSA 2015 (7th Edition: Cebu, December 2015), CSA 2014 (6th Edition: Guam, December, 2014), CSA 2013 (5th Edition: Danang, December, 2013), CSA 2012 (4th Edition: Jeju, November, 2012), CSA 2011 (3rd Edition: Jeju, December, 2011), CSA 2009 (2nd Edition: Jeju, December, 2009), and CSA 2008 (1st Edition: Australia, October, 2008).

CSA 2016 contains high-quality research papers submitted by researchers from all over the world. Each submitted paper was peer reviewed by reviewers who are experts in the subject area of the paper. Based on the review results, the program committee accepted papers.

For organizing an international conference, the support and help of many people is needed. First, we would like to thank all authors for submitting their papers. We also appreciate the support from program committee members and reviewers who carried out the most difficult work of carefully evaluating the submitted papers.

We would like to give my special thanks to Prof. James J. (Jong Hyuk) Park, Prof. Han-Chieh Chao, and Prof. Mohammad S. Obaidat, the Steering Committee Chairs of CSA for their strong encouragement and guidance to organize the symposium. We would like to thank CSA 2016 General Chair, Prof. Young-Sik Jeong, Prof. Yang Xiao, Prof. Vincenzo Loia, Prof. Rung-Shiang Cheng, Prof. Victor Leung, and

Prof. Ka Lok Man. We would like to express special thanks to committee members for their timely unlimited support.

CSA 2016 Program Chairs
Gangman Yi
Houcine Hassan
Yu Chen
Hwamin Lee
Jin Wang

Organization

Honorary Chair

Doo-soon Park	SoonChunHyang University/KIPS President, Korea

Steering Chairs

James J. Park	SeoulTech, Korea
Han-Chieh Chao	National Ilan University, Taiwan
Mohammad S. Obaidat	Monmouth University, USA

General Chairs

Young-Sik Jeong	Dongguk University, Korea
Yang Xiao	University of Alabama, USA
Vincenzo Loia	University of Salerno, Italy
Rung-Shiang Cheng	Kunshan University, Taiwan
Victor Leung	University of British Columbia, Canada
Ka Lok Man	Xi'an Jiaotong-Liverpool University, China

Program Chairs

Gangman Yi	Dongguk University, Korea
Houcine Hassan	Universitat Politècnica de València, Spain
Yu Chen	State University of New York, USA
Hwamin Lee	Soonchunhyang University, Korea
Jin Wang	Yangzhou University, China

Program Vice-chair

Yunsick Sung	Keimyung University, Korea

Workshop Chairs

Neil Y. Yen	The University of Aizu, Japan
Kyungeun Cho	Dongguk University, Korea
Se Dong Min	SoonChunHyang University, Korea

International Advisory Board

Simon James Fong	University of Macau, Macau, China
Yi Pan	Georgia State University, USA
Michael Hwa Young Jeong	Kyung Hee University, Korea
Hsiao-Hwa Chen	Sun Yat-Sen University, Taiwan
Philip S. Yu	University of Illinois at Chicago, USA
Jiankun Hu	RMIT University, Australia
Shu-Ching Chen	Florida International University, USA
Victor Leung	University of British Columbia, Canada
Qun Jin	Waseda University, Japan
Frode Eika Sandnes	Oslo University College, Norway
Stephan Olariu	Old Dominion University, USA
Koji Nakano	University of Hiroshima, Japan

Publicity Chairs

Mohammad Maifi Hasan Khan	University of Connecticut, USA
Myeonggil Choi	Chung Ang University, Korea
Jungho Kang	Sungsil University, Korea
Seokhoon Kim	Soonchunhyang University, Korea
Haixia Zhang	Shandong University, China
Rafael Falcon	Larus Technologies, Canada
Xu Shao	Institute for Infocomm Research, Singapore
Amiya Nayak	University of Ottawa, Canada
Zhen Huang	University of Ottawa, Canada
Weiwei Fang	Beijing Jiaotong University, China
Kevin Cheng	Tatung University, Taiwan

Program Committee

Andrew Kusiak	The University of Iowa, USA
Ardagna Claudio	University of Milan, Italy
Basarir Metin	Sakarya University, Turkey
Bela Genge	University of Targu Mures, Romania
Chang Wu Yu	Chung Hua University, Taiwan
Chia-Hung Yeh	National Sun Yat-sen University, Taiwan
Chin-Fu Kuo	The National Kaohsiung University, Taiwan
Cho-Chin Lin	National Yilan University, Taiwan
Chuan-Ming Liu	National Taipei University of Technology, Taipei
Debajyoti Mukhopadhyay	Balaji Institute of Telecom & Management, India
Dion Hoe-Lian Goh	Nanyang Technological University, Singapore
El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Eunyoung Lee	Dongduk University, Korea
Guan-Ling Lee	National Dong Hwa University, Taiwan
Jae Joon Lee	Ajou University, Korea
Jehn-Ruey Jiang	National Central University, Taiwan
Jerzy Respondek	Silesian University of Technology Poland
Kenny Adamson	University of Ulster, UK
Kuei-Ping Shih	Tamkang University, Taiwan
Listanti Marco	DIET, Roma, Italy
M. Dominguez Morales	University of Seville, Spain
Massimo Cafaro	University of Salento, Italy
Muhammad Javed	Dublin City University, Ireland
Paprzycki Marcin	Polish Academy of Sciences, Poland
Qian Yu	University of Regina, Canada
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Alok Desai	Brigham Young University, USA
Ana Isabel Pereira	Polytechnic Institute of Braganca, Portugal

Message from the CUTE 2016 General Chairs

On behalf of the organizing committees, it is our pleasure to welcome you to the 11th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2016), will be held in Bangkok, Thailand, duringn December 19–21, 2016.

This conference provides an international forum for the presentation and showcase of recent advances on various aspects of ubiquitous computing. It will reflect the state of the art of the computational methods, involving theory, algorithm, numerical simulation, error and uncertainty analysis and/or novel application of new processing techniques in engineering, science, and other disciplines related to ubiquitous computing.

The papers included in the proceedings cover the following topics: Ubiquitous communication and networking, Ubiquitous software technology, Ubiquitous systems and applications, Ubiquitous security, Privacy and trust. Accepted papers highlight new trends and challenges in the field of ubiquitous computing technologies. We hope you will find these results useful and inspiring for your future research.

We would like to express our sincere thanks to General Vice-Chairs: James J. Park (SeoulTech, Korea), Young-Sik Jeong (Dongguk University, Korea), Doo-soon Park (SoonChunHyang University, Korea), Laurence T. Yang (St. Francis Xavier University, Canada), Hai Jin (Huangzhong University of Science and Technology, China), Chan-Hyun Youn (KAIST, Korea), Jianhua Ma (Hosei University, Japan), Mingyi Guo (Shanghai Jiao Tong University, China), and Weijia Jia (City University of Hong Kong, Hong Kong). We would also like to express our cordial thanks to the Program Committee members for their valuable efforts in the review process, which helped us to guarantee the highest quality of the selected papers for the conference.

Finally, we would thank all the authors for their valuable contributions and the other participants of this conference. The conference would not have been possible without their support. Thanks are also due to the many experts who contributed to making the event a success.

CUTE 2016 General Chairs
Yi Pan
Weimin Zheng
Gangman Yi
Myung-Hyun Yoon
No Byung-Gyu
Hyoung Woo Park

Message from the CUTE 2016

Program Chairs

Welcome to the 10th International Conference on Ubiquitous Information Technologies and Applications (CUTE 2016), which will be held in Bangkok, Thailand during December 19–21, 2016.

The purpose of the CUTE 2016 conference is to promote discussion and interaction among academics, researchers and professionals in the field of ubiquitous computing technologies. This year the value, breadth, and depth of the CUTE 2016 conference continues to strengthen and grow in importance for both the academic and industrial communities. This strength is evidenced this year by having the highest number of submissions made to the conference.

For CUTE 2016, we received a lot of paper submissions from various countries. Out of these, after a rigorous peer review process, we accepted only high-quality papers for CUTE 2016 proceedings, published by the Springer. All submitted papers have undergone blind reviews by at least two reviewers from the technical program committee, which consists of leading researchers around the globe. Without their hard work, achieving such a high-quality proceeding would not have been possible. We take this opportunity to thank them for their great support and cooperation.

Finally, we would like to thank all of you for your participation in our conference, and also thank all the authors, reviewers, and organizing committee members. Thank you and enjoy the conference!

CUTE 2016 Program Chairs
Yunsick Sung
Keqiu Li
Eunyoung Lee

Organization

Honorary Chair

Doo-soon Park	SoonChunHyang University, Korea

Steering Committee

James J. Park	SeoulTech, Korea (Leading Chair)
Young-Sik Jeong	Dongguk University, Korea (Co-chair)
Doo-soon Park	SoonChunHyang University, Korea (Co-chair)
Laurence T. Yang	St. Francis Xavier University, Canada
Hai Jin	Huangzhong University of Science and Technology, China
Chan-Hyun Youn	KAIST, Korea
Jianhua Ma	Hosei University, Japan
Mingyi Guo	Shanghai Jiao Tong University, China
Weijia Jia	City University of Hong Kong, Hong Kong

General Chairs

Yi Pan	Georgia State University, USA
Weimin Zheng	Tsinghua University, China
Gangman Yi	Dongguk University, Korea
Myung-Hyun Yoon	KETI, Korea
No Byung-Gyu	KISA, Korea
Hyoung Woo Park	KISTI, Korea

Program Chairs

Yunsick Sung	Keimyung University, Korea
Keqiu Li	Dalian University of Technology, China
Eunyoung Lee	Dongduk Women's University, Korea

Program Vice-Chairs (PVC)

Jason C. Hung	Oversea Chinese University, Taiwan
Shuhui Yang	Purdue University, USA
Xuan Guo	Oak Ridge National Lab, USA
Kwang-il Hwang	Incheon National University, Korea
Lei Ye	University of Wollongong, Australia
KwangMan Ko	Sangji University, Korea
Antonio Coronato	ICAR-CNR, Italy
Yoo-Joo Choi	Seoul Media Institute of Technology, Korea
Dae-Sung Moon	ETRI, Korea
Chang-Sun Shin	Sunchon National University, Korea
Shingchern You	National Taipei University of Technology, Taiwan
Okyeon Yi	Kookmin University, Korea
Dongwon Jeong	Kunsan National University, Korea
Soo-Hyun Park	Kookmin University, Korea
Sung Bum Pan	Chosun University, Korea
Min Choi	Chungbuk National University, Korea
Joon-Min Gil	Catholic University of Daegu, Korea
Namje Park	Jeju National University, Korea
Jin Ho Yoo	Sangmyung University, Korea
Deok Gyu Lee	Seowon University, Korea
Yong-Ik Yoon	Sookmyung Women's University, Korea

Workshop Chairs

Young-Gab Kim	Sejong University, Korea
Eunmi Choi	Kookmin University, Korea
Aziz Nasridinov	Chungbuk National University, South Korea

International Advisory Committee

Seok Cheon Park	Gachon University, Korea
Im-Yeong Lee	SoonChunHyang University, Korea
Sanghoon Kim	Hankyong National University, Korea
HeonChang Yu	Korea University, Korea
Nammee Moon	Hoseo University, Korea
Byeong-Seok Shin	Inha, Korea
Dong-Ho Kim	Soongsil, Korea
Keun Ho Ryu	Chungbuk National University, Korea
JaeKwang Lee	Hannam University, Korea
Yoo-jae Won	Chungnam National University, Korea

Publicity Chairs

Min Li	Central South University, China
Junbo Zhang	Microsoft Research, Beijing, China
Seung-Won Jung	Dongguk University, Korea
Jong-Hyouk Lee	Sangmyung University, Korea
Jaehwa Chung	Korea National Open University, Korea

Industrial Workshop Chairs

Jungho Kang	Soongsil University, Korea
Sung Chul Yu	LG Hitachi Co. Ltd., Korea

Program Committee

Abdullah Gani	Dept. of Computer System & Technology, University of Malaya, Malaysia
Rosil Salleh	Dept. of Computer System & Technology, University of Malaya, Malaysia
Siti Hafizah	Dept. of Computer System & Technology, University of Malaya, Malaysia
Anjum Naveed	Dept. of Computer System & Technology, University of Malaya, Malaysia
Ejaz Ahmed	Centre for Mobile Cloud Computing Research (C4MCCR), Pakistan
Jonghee Youn	Dept. of Computer Engineering, Yeungnam University, Korea
Doo-San Cho	Dept. of Electronics Engineering, Sunchon National University, Korea
Hee-Wa Park	Dept. of Information & Communication, Halla University, Korea
Chang Won Jeong	School of Electrical, Electronic and Information Engineering, Wonkwang University, Korea
YunHee Kang	Division of Information & Communication, Baekseok University, Korea
Bong-Kee Sin	Dept. of IT Convergence and Applications Engineering, Pukyong National University, Korea
Mun-Kyu Lee	Dept. of Computer and Information Engineering, Inha University, Korea
Jong-Min Kim	Dept. of IT Management, Kosin University, Korea

Sungyoung Kim	Dept. of Computer Engineering, Kumoh National Institute of Technology, Korea
Jaepil Ko	Dept. of Computer Engineering, Kumoh National Institute of Technology, Korea
Byoungwoo Oh	Dept. of Computer Engineering, Kumoh National Institute of Technology, Korea
Kyung-Soo Lim	Hyper-connected Communication Research Laboratory, ETRI, Korea
Hansung Lee	SW R&D Center Samsung Electronics, Korea
Hoonju Chung	School of Electronic Engineering, Kumoh National Institute of Technology, Korea
Yonghwan Lee	School of Electronic Engineering, Kumoh National Institute of Technology, Korea
Yong-Seok Kim	Department of Computer Information, Seonam University, Korea
Hoon-Jae Lee	Division of Computer Eng., Dongseo University, Korea
Jun-Won Ho	Department of Information Security, Seoul Women's University, Korea
Jang Ho Lee	Dept. of Computer Engineering, Hongik University, Korea
Youngsun Han	Department of Electronic Engineering, Kyungil University, Korea
Kwang Sik Chung	Dept. of Computer Science, Korea National Open University, Korea
Neungsoo Park	Dept. of Computer Scienece & Engineering, Konkuk University, Korea
Sangheon Pack	School of Electrical Engineering, Korea University, Korea
Yun-hee Kang	Division of Information and Communication, Baekseok University, Korea
Seong Gon Choi	School of Information and Communication Engineering, Chungbuk National University, Korea
Heewan Park	School of Information Communication & Broadcasting Engineering, Halla University, Korea
HwaMin Lee	Dept. of Computer Software Engineering, Soonchunhyang University, Korea
Yongyun Cho	Information and Communication Engineering, Sunchon National University, Korea
Ki-Sik Kong	Dept. of Multimedia, Namseoul University, Korea
Keun Chang Kwak	Dept. of Electronics Engineering, Chosun University, Korea

ByungRae Cha	School of Information & Communications, GIST, Korea
Sun Park	School of Information & Communications, GIST, Korea
Binod Vaidya	Assoc. Director at BroadWIRLab, EECS, University of Ottawa, Canada
Nishat Ahmad	Pakistan University, Pakistan
Yongwha Chung	Korea University, Korea
Wooyong Choi	ETRI, Korea
Seunghoon Chae	ETRI, Korea
Haemin Moon	Chosun University, Korea
Sungsuk Kim	Dept. of Computer Science, Seokyeong University, Korea
Dong-Mahn Seo	School of Information Technology Eng., Catholic University of Daegu, Korea
Dong-Kyun Kim	KISTI, Korea
Jin-Hyung Park	KISTI, Korea
Sang-Sun Byun	Catholic University of Pusan, Korea
Jong-Bum Lim	Dongguk University, Korea
Youn-Hee Han	School of Computer Science and Engineering, Korea University of Technology and Education, Korea
Buseong Cho	Department of KREONET Operation and Service, KISTI, Korea
Woojin Seok	Advanced KREONET Application and Support Team, KISTI, Korea
Sung-Hwa Hong	Dept of Maritime Inform. & Comm., Mokpo National Martitime University, Korea
Ji Su Park	National Center of Excellence in Software, Chungnam National University, Korea
Chan Yeol Park	KISTI, Korea
Jangwon Choi	KISTI, Korea
Heongchang Yu	Dept. of Computer Science and Engineering, Korea University, Korea
Myoung Sun Noh	KISA, Korea
Won Tae Sim	KISA, Korea
Dong Young Yoo	KISA, Korea
Sangsoo Jang	KISA, Korea
Jonghyun Baek	KISA, Korea
Yonghwan Kim	KEIT, Korea
Jongho Baek	Seoul Women's University, Korea
Jongwan Kim	Department of Computer Engineering, Sungkyul University, Korea
Jong-Won Lee	Dept. of Electrical and Computer Engineering, Ajou University, Korea

Jong-Kook Kim	School of Electrical Engineering, Korea University, Korea
Bo-Chao Cheng	National Chung-Cheng University, Taiwan
Chang Yao-Chung	National Taitung University, Taiwan
Dumitru Roman	SINTEF/University of Oslo, Norway
Eunmi Choi	Kookmin University, Korea
Heonchang Yu	Korea University, Korea
Imad Saleh	University of Paris 8, France
Jeong-Hyon Hwang	State University of New York at Albany, USA
Jiang Fuu-Cheng	Tunghai University, Taiwan
Jin-Hee Cho	U.S. Army Research Laboratory, USA
Jong-Myon Kim	University of Ulsan, Korea
Kwang Sik Chung	Korea National Open University, Korea
Q. Shi	Liverpool John Moores University, UK
Wei Wei	Xi'an University of Technology, China
Wen-Chi Hou	Southern Illinois University, USA
Yang-Sae Moon	Kangwon National University, Korea

Contents

Advances in Information Technologies and Applications

Design and Development of a Robotic Arm for Rehabilitation and Training 3
Sarut Panjan and Siam Charoenseang

Detection of a Robust High-Frequency Range via Noise Analysis in a Real-World Environment 9
Myoungbeom Chung and Ilju Ko

Intelligent Food Distribution Monitoring System 16
Ganjar Alfian, Hyejung Ahn, Yoonmo Shin, Jaeho Lee and Jongtae Rhee

Design of Sudden Unintended Acceleration Check System Using Distance Measurement Sensor 23
Jea-Hui Cha, Tae-Hyoung Kim and Jong-Wook Jang

Real-Time Dynamic Motion Capture Using Multiple Kinects 29
Seongmin Baek and Myunggyu Kim

Cell-Based Indexing Method for Spatial Data Management in Hybrid Cloud Systems 36
Yan Li and Byeong-Seok Shin

SOA Based Equipment Data Management System for Smart Factory . . . 42
YunHee Kang, Soong-ho Ko and Kyoungwoo Kang

The Problem Analysis of Specific Personal Information Protection Assessment in Japan Case 48
Sanggyu Shin, Yoichi Seto, Mayumi Sasaki and Kei Sakamoto

Using a Fine-Grained Hybrid Feature for Malware Similarity Analysis 54
Jing Liu, Yongjun Wang, Peidai Xie and Xingkong Ma

A Wireless Kinect Sensor Network System for Virtual Reality Applications 61
Mengxuan Li, Wei Song, Liang Song, Kaisi Huang, Yulong Xi and Kyungeun Cho

Finding Comfortable Settings of Snake Game Using Game Refinement Measurement 66
Anunpattana Punyawee, Chetprayoon Panumate and Hiroyuki Iida

Code Modification and Obfuscation Detection Test Using Malicious Script Distributing Website Inspection Technology 74
Seong-Min Park, Han-Chul Bae, Young-Tae Cha and Hwan-Kuk Kim

Initialization of Software Defined Wireless Bacteria-Inspired Network Platform 81
Shih-Yun Huang, Hsin-Hung Cho, Yu-Zen Wang, Timothy K. Shih and Han-Chieh Chao

An Extension of QSL for E-voting Systems 87
Yuan Zhou, Hongbiao Gao and Jingde Cheng

Behavior-Based Detection for Malicious Script-Based Attack 97
Soojin Yoon, Hyun-lock Choo, Hanchul Bae and Hwankuk Kim

The SP-tree: A Clustered Index Structure for Efficient Sequential Access 104
Guang-Ho Cha

An Address Conflict Resolving Scheme of Inter-drone Ad Hoc Communications for Hide Densely Deployed Low Power Wide Area Networks 111
Jaeho Lee and Bong-Ki Son

State-of-the-Art Algorithms for Mining Up-to-Date High Average-Utility Patterns 117
Donggyu Kim and Unil Yun

Design of Shoot'em up Game Using OpenGL 122
Unil Yun and Heungmo Ryang

Performance Analysis of Tree-Based Algorithms for Incremental High Utility Pattern Mining 127
Heungmo Ryang and Unil Yun

Development of 2D Side-Scrolling Running Game Using the Unity 3D Game Engine 132
Wooseong Jeong and Unil Yun

EPD Noticeboard for Posting Multiple Information 137
Bong-Ki Son and Jaeho Lee

Design of Processing Model for Connected Car Data Using Big Data Technology 143
Lionel Nkenyereye and Jong Wook Jang

Efficient Path Selection for IoT Devices in Heterogeneous Service Environments 149
Dae-Young Kim and Seokhoon Kim

Forecasting Sugarcane Yield Using (μ+λ) Adaptive Evolution Strategies 154
Supawadee Srikamdee, Sunisa Rimcharoen and Nutthanon Leelathakul

Resource Pooling Mechanism for Mobile Cloud Computing Service . . . 160
Seok-Hyeon Han, Hyun-Woo Kim and Young-Sik Jeong

IPC Multi-label Classification Applying the Characteristics of Patent Documents 166
Sora Lim and YongJin Kwon

A Comparison of Data Mining Methods in Analyzing Educational Data 173
Euihyun Jung

A New Secure Android Model Based on Privilege 179
Tao Zhang and Zhilong Wang

Survey of MCC Architectures for Computing Service 185
Byeong-Seok Park, Yoon-A Heo and Young-Sik Jeong

Measurement of Enterprise Smart Business Capability in a Global Management Environment 192
Chui Young Yoon

Occluded Pedestrian Classification Using Gradient Patch and Convolutional Neural Networks 198
Sangyoon Kim and Moonhyun Kim

A Design of Secure Authentication Method with Bio-Information in the Car Sharing Environment 205
Sang-Hyeon Park, Jeong-Ho Kim and Moon-Seog Jun

A Design of Certificateless-Based Device Authentication Scheme in the SmartHome Environment 211
Jae Seung Lee, Jaehwa Chung and Sangkee Suk

A Design of Secure Authentication Method Using Zero Knowledge Proof in Smart-Home Environment 215
Geunil Park, Bumryoung Kim and Moon-seog Jun

Drone Classification by Available Control Distances 221
Mansik Kim, Hyungjoo Kim, Jungho Kang and Jaesoo Kim

A Design of Key Agreement Scheme Between Lightweight Devices in IoT Environment 224
Hague-Chung, Keun-Chang Choi and Moon-Seog Jun

Platform Independent Workflow Mechanism for Bigdata Analytics 230
Tai-Yeon Ku, Hee-Sun Won and Hoon Choi

Water Surface Simulation Based on Perlin Noise and Secondary Distorted Textures 236
Hua Li, Huamin Yang, Chao Xu and Yuling Cao

Optimization for Particle Filter-Based Object Tracking in Embedded Systems Using Parallel Programming 246
Mai Thanh Nhat Truong and Sanghoon Kim

Generalized Multi-linear Mixed Effects Model 253
Chao Li, Lili Guo, Zheng Dou, Guangzhen Si and Chunmei Li

Anomaly Detection of Spectrum in Wireless Communication via Deep Autoencoder 259
Qingsong Feng, Zheng Dou, Chunmei Li and Guangzhen Si

A Modified Complex ICA for Blind Source Separation and the Application in Communication Reconnaissance 266
Zheng Dou, Zi Xiao, Yang Zhao and Jinyu Wang

Sensor Coverage Problem in Sparse MANET Environments 273
JongBeom Lim, HeonChang Yu and JoonMin Gil

Design of Jitter Buffer Control Algorithm for Guaranteeing the Medical Information Data Transmission Quality in Wireless Network Environment 279
Dong-Wan Joe, Jae-Sung Shim, Yong-Wan Ju and Seok-Cheon Park

Student's-t Mixture Model Based Excepted Patch Log Likelihood Method for Image Denoising 285
J.W. Zhang, J. Liu, Y.H. Zheng and J. Wang

Regularization Parameter Selection for Gaussian Mixture Model Based Image Denoising Method 291
J.W. Zhang, J. Liu, Y.H. Zheng and J. Wang

Restoration Method for Satellite Image Based on Content-Aware Reciprocal Cell Pool 298
Yuhui Zheng, Xiaozhou Zhou, Tong Li and Jin Wang

Student's-t Mixture Model Based Image Denoising Method with Gradient Fidelity Term 306
J.W. Zhang, J. Liu, Y.H. Zheng and J. Wang

Classification Algorithms for Privacy Preserving in Data Mining: A Survey 312
Sai Ji, Zhen Wang, Qi Liu and Xiaodong Liu

Exploiting Group Signature to Implement User Authentication in Cloud Computing 323
Sai Ji, Dengzhi Liu and Jian Shen

A Secure System Framework for an Agricultural IoT Application 332
Hao Wu, Fangpeng Chen, Hanfeng Hu, Qi Liu and Sai Ji

Localization Technology in Wireless Sensor Networks Using RSSI and LQI: A Survey 342
Sai Ji, Dengzhi Liu and Jian Shen

Energy-Balanced Unequal Clustering Routing Algorithm for Wireless Sensor Networks 352
Jin Wang, Yiquan Cao, Jiayi Cao, Huan Ji and Xiaofeng Yu

The Agent Communication Simulation Based on the Ego State Model of Transactional Analysis 360
Mi-Sun Kim, Green Bang and Il-Ju Ko

Design of Software Reliability Test Architecture for the Connected Car 366
Doo-soon Park and Seokhoon Kim

Network Activation Control According to Traffic Characteristics in Sensor Networks for IoT 371
Dae-Young Kim, Young-Sik Jeong and Seokhoon Kim

Data Mining Techniques to Facilitate Digital Forensics Investigations . . . 376
Erik Miranda Lopez, Yoon Ho Kim and Jong Hyuk Park

Solving Standard Cell Placement Problem Using Discrete Firefly Algorithm: A Nature Inspired Approach 380
Pradip Kumar Sharma, Saurabh Singh and Jong Hyuk Park

A Security Model for Protecting Virtualization in Cloud Computing 385
Saurabh Singh, Pradip Kumar Sharma and Jong Hyuk Park

Effective Pre-processing Methods with DTG Big Data by Using MapReduce Techniques 389
Wonhee Cho and Eunmi Choi

Security Requirements and Countermeasures for Secure Home Network in Internet of Things 396
Seo Yeon Moon, Saurabh Singh and Jong Hyuk Park

Machine Learning for Trajectory Generation of Multiple-pedestrians . . . 401
Hye-Yeon Yu, Young-Nam Kim and Moon-Hyun Kim

A Deep Learning-Based Gait Posture Recognition from Depth Information for Smart Home Applications 407
Md. Zia Uddin and Mi Ryang Kim

Implementation of an Image Restoration with Block Iteration Method for Spatially Variant Blur Models 414
Ji Yeon Lee, Kuk Won Ko and Sangjoon Lee

Analysis of Hard Shadow Anti-aliasing 421
Hua Li, Yuling Cao and Xin Feng

Cascade-Adaboost for Pedestrian Detection Using HOG and Combined Features 430
Gyujin Jang, Jinhee Park and Moonhyun Kim

Portable Hypervisor Design for Commercial 64-Bit Android Devices Supporting 32-Bit Compatible Mode 436
Kangho Kim, Kwangwon Koh, Seunghyub Jeon and Sungin Jung

Concept-Based Compound Keyword Extraction Based on Using Sentential Distance, Conceptual Distance and Production Rules: Calculation of the Keyword Importance 442
Samuel Sangkon Lee

Coarse-Grained 2.5-D CSAMT Parallel Inversion Method Based on Multi-core CPU 449
Lili He, Hongtao Bai, Jin Wang, Yu Jiang and Tonglin Li

Virtual Force and Glowworm Swarm Optimization Based Node Deployment Strategy for WSNs 456
Jin Wang, Yiquan Cao, Jiayi Cao, Huan Ji and Xiaofeng Yu

Comparison with Recommendation Algorithm Based on Random Forest Model 463
Yu Jiang, Lili He, Yan Gao, Kai Wang and Chengquan Hu

Risk Factors and the Difference Among Hypertension, Diabetes and Heart Disease 471
Xue Wang, Lili He and Hongtao Bai

Linear Programming Computation Model Based on DPVM 477
Hongtao Bai, Lili He, Yu Jiang, Jin Wang and Shanshan Jiang

A Study on the Effective Communication Protocol of the Surface Inspection Rail Robot that it can be a Self-checking 484
Yun-Seok Lee, Eun Kim, Sungyun Kim and Seokhoon Kim

The Efficient Multimedia Transmission Services for the E-learning System with Sensor 490
Sung-Hwa Hong and Joon-Min Gil

A Quality Model for IoT Service 497
Mi Kim, Jin Ho Park and Nam Yong Lee

An Empirical Study of Risk Factors for the Development Methodology for Small-Size IT Projects 505
Joon Ho Park, Nam Yong Lee and Jin Ho Park

Nack-Based Broadcast Mechanism for Isochronous Audio Stream Transmission Using Bluetooth Low Energy 511
Jaeho Lee

Empirical Study on IoT-Learning for the Rehabilitation Treatment of Chronic Low Back Pain Patients 517
Seul-Ah Shin, Ji-Soo Choi, Young-Jong Kim, Nam-Yong Lee and Jin-Ho Park

A Theoretical Study of Hardware Architecture for Network Security Server 525
Joong-Yeon Lee, In-Taek Oh, Nam-Yong Lee and Jin-Ho Park

An Empirical Study of the Relationship Between DISC Behavioral Style of Application Programmer and Quality of Software Development 531
In-Taek Oh, Joong-Yeon Lee, Jae-Yoon Cheon, Nam-yong Lee and Jin-Ho Park

Advances in Computer Science and Ubiquitous Computing

The Development of COB Type LED Lighting System for High Temperature Machine Vision 543
Park Sanggug

An Alternative Management Scheme of DHCP Lease Time for Internet of Things 549
Pyung Soo Kim, Eung Hyuk Lee and Eung Tae Kim

A User Empirical Context Model for a Smart Home Simulator 555
Green Bang and Ilju Ko

Co-display Content Service for First-Person Videos of Smart Glass 561
Bokyung Sung and Ilju Ko

Probabilistic Analysis for the Relationship Between Min-Entropy and Guessing Attack 567
Ju-Sung Kang, Hojoong Park and Yongjin Yeom

Dynamic QoS Scheme for InfiniBand-Based Clusters 573
Bongjae Kim and Jeong-Dong Kim

Applying PE-Miner Framework to Software Defined Network Quarantine 579
Dong-Hee Kim, Soo-Hwan Lee, Won-Sik Doo, Sang-Il Ahn and Tai-Myoung Chung

A Novel Method for Eliminating Duplicated Frames in Ethernet Standard (IEEE 802.3) Networks 585
Saad Allawi Nsaif and Jong Myung Rhee

A Study of Malicious Code Classification System Using MinHash in Network Quarantine Using SDN 594
Soo-Hwan Lee, Myeong-Uk Song, Jun-Kwon Jung and Tai-Myoung Chung

Application of RFID and Computer Vision for the Inventory Management System 600
Ganjar Alfian, Jaeho Lee, Hyejung Ahn and Jongtae Rhee

Prediction Method for Suspicious Behavior Based on Omni-View Model 607
Ji-Hyen Choi, Jong-Won Choe and Yong-Ik Yoon

Optimal 3D Printing Direction for Stability of Slanted Shapes 613
Jiyoung Park and Hwa Seon Shin

A Study on DDS-Based BLE Profile Adaptor for Solving BLE Data Heterogeneity in Internet of Things 619
Jung-Hoon Oh, Moon-Ki Back, Gil-Tak Oh and Kyu-Chul Lee

A Study of Environment-Adaptive Intrusion Detection System 625
Ki-Hyun Lee and Young B. Park

OFART: OpenFlow-Switch Adaptive Random Testing 631
Dong-Su Koo and Young B. Park

An Evaluation of Availability, Reliability and Power Consumption for a SDN Infrastructure Using Stochastic Reward Net 637
Kihong Han, Tuan Anh Nguyen, Dugki Min and Eun Mi Choi

A Dynamic Service Binding Scheme with Service OID for IoT 649
Euihyun Jung

A Context-Aware Architecture Pattern to Enhance the Flexibility of Software Artifacts Reuse 654
Doohwan Kim, Soon-Kyeom Kim, Woosung Jung and Jang-Eui Hong

Deep Analysis of Tag Interference by Tag to Tag Relative Angles with Passive Far Field UHF RFID System 660
Jae Sung Choi and Hyun Lee

Simple Method of Video Mapping of Multiple Targets 665
In-Jae Jo, Joohun Lee and Yoo-Joo Choi

Evolutionary Test Case Generation from UML-Diagram with Concurrency 674
Seungchan Back, Hyorin Choi, Jung-Won Lee and Byungjeong Lee

Evaluating the Effectiveness of the Vector Space Retrieval Model Indexing 680
Jung-Hoon Shin, Mesfin Abebe, Cheol Jung Yoo, Suntae Kim, Jeong Hyu Lee and Hee-Kyung Yoo

Active Tracking Strategy with Multiple Cameras in Large Areas 686
Sangjin Hong and Nammee Moon

A Survey and Design of a Scalable Mobile Edge Cloud Platform for the Smart IoT Devices and It's Applications 694
Yeongpil Cho, Yunheung Paek, Ejaz Ahmed and Kwangman Ko

Network Anomaly Detection Based on Probabilistic Analysis 699
JinSoo Park, Dong Hag Choi, You-Boo Jeon, Se Dong Min and Doo-Soon Park

Hedonic Model Study for Retargeting Advertising Based on Space-Centered Internet of Things 705
Bo-Ram Kim, Man-Soo Chung and Yong-Ik Yoon

A New Automated Cell Counting Program by Using Hough Transform-Based Double Edge 712
Jae Sung Choi, Moon Jong Choi, Jung-Min Lee and Hyun Lee

An Approach for Interworking Heterogeneous Networks with DTN and IP Routing in Space Internet 717
Euiri An, Kyungrak Lee, Jaewon Lee and Inwhee Joe

Implementation of Recommender System Based on Personalized Search Using Intimacy in SNS 723
Jeong-Dong Kim, Bongjae Kim and Jeong-Ho Park

Measuring Similarity Between Graphs Based on Formal Concept Analysis 730
Fei Hao, Dae-Soo Sim and Doo-Soon Park

TEXAS2: A System for Extracting Domain Topic Using Link Analysis and Searching for Relevant Features 736
SangWon Hwang, YongSeok Lee and YoungKwang Nam

Ubiquitous Computing for Cloud Infrastructure to Mobile Application in IoT Environment 742
DongBum Seo, Keun-Ho Lee and You-Boo Jeon

What Are Learning Satisfaction Factors in Flipped Learning? 750
Kyung Yeul Kim and Yong Kim

Development of UI Guideline for Senior Citizens' e-Learning Content 756
Myung In Kim and Yong Kim

Full Duplex Relaying with Buffer Based on Cognitive Radio Technique 762
Junsu Kim, Doo-Hee Jung, Jeho Lee and Su Min Kim

Design of Docking Drone System Using P-PID Flight Controller 768
Beck Jong-Hwan, Pak Myeong-Suk and Kim Sang-Hoon

Lightweight Security for Underwater IoT 774
Sun-Ho Yeom, Jung-Il Namgung, Soo-Young Shin and Soo-Hyun Park

Image Based Video Querying Algorithm Using 3-Level Haar Wavelet Transform Features 779
Changseok Bae, Yuk Ying Chung and Jeunwoo Lee

Design and Implementation of the Mobile Learning App for Creative Problem Solving Activities 786
Ji-Hye Bae and Hyun Lee

IoT-Based Smart Photo Frame 793
Ji-Hye Bae, In-Hwan Kim, Yong-Tae Jeon and Hyun Lee

A Fast Algorithm for Generating Virtual Dedicate Network Based on Software-Defined Wide Area Network 799
Yong-hwan Kim, Buseung Cho and Dongkyun Kim

Detection of Content Changes Based on Deep Neural Networks 806
Noo-ri Kim, YunSeok Choi, HyunSoo Lee and Jee-Hyong Lee

Resource Allocation in D2D Networks with Location Based Distance Information 812
Soo Hyeong Kang, Pyung Soo Kim, Bang Won Seo and Jeong Gon Kim

Design of Corporate Business Card Management System 819
Seok-heon Ko, Gil-mo Yang and Jun-dong Lee

The OpenWRT's Random Number Generator Designed Like /dev/urandom and Its Vulnerability 825
Dongchang Yoo and Yongjin Yeom

Implementation of the Block2 Option Transfer for Resource Observing with the CoAPthon Library 831
Kyoung-Han Kim, Hyun-Kyo Lim, Joo-Seong Heo and Youn-Hee Han

Hive-Based Anomaly Detection in Hadoop Log Data Management 837
Siwoon Son, Myeong-Seon Gil, Seokwoo Yang and Yang-Sae Moon

HIM-PRS: A Patent Recommendation System Based on Hierarchical Index-Based MapReduce Framework 843
Xuhua Rui and Dugki Min

Finding Meaningful Chronological Pattern of Key Words in Computer Science Bibliography 849
Joo-Seong Heo, Hyun-Kyo Lim, Kyong-Han Kim and Youn-Hee Han

The Design of Intelligent Video Analytics System Performing Automatic Noise Rejection by Comparing Distribution of Metadata of Moving Object 855
Taewoo Kim, Hyungheon Kim and Pyeongkang Kim

Dependability Analysis of Digital Library Cloud Services with Load Sharing 861
Dongseok Lee, Sungsoo Kim and Tae-Sun Chung

Document Classification Using Word2Vec and Chi-square on Apache Spark 867
Mijin Choi, Rize Jin and Tae-Sun Chung

Analysis of Recent Maximal Frequent Pattern Mining Approaches 873
Gangin Lee and Unil Yun

Design of Noise Information Storage System Using IoT Devices 878
Judae Lee and Unil Yun

Analysis of Privacy Preserving Approaches in High Utility Pattern Mining 883
Unil Yun and Donggyu Kim

An Agent-Based Remote Operation and Safety Monitoring System for Marine Elevators 888
Hyung-Joo Kim and Kwangil Lee

Survey on CPN Applications in Cloud Computing 894
Rustam Rakhimov Igorevich and Dugki Min

An Observation Method for Estimating Carrier Frequency Offset in OFDM Systems 900
Mustafa Altaha and Humor Hwang

Korean-to-Korean Translation Based Learning Contents Management System for Parents of Multi-cultural Family 906
YunHee Kang, Myung Ju Kang and WooSik Kim

A Distributed Survey Automation Based on a Customizable Form Template 912
Jaekwon Lee, Kisub Kim, Jang-Eui Hong and Woosung Jung

Mobile Agent Oriented Service for Offloading on Mobile Cloud Computing 920
HwiRim Byun, Boo-Kwang Park and Young-Sik Jeong

Unstructured Data Service Model Utilizing Context-Aware Big Data Analysis 926
Yonghoon Kim and Mokdong Chung

Information Reminder System Based on Word Registered by User 932
KyeYoung Kim, Byeong-Eon Ahn, Suk-Young Lim, Daejin Moon and Dae-Soo Cho

A Study of Determining Abnormal Behaviors by Using System for Preventing Agricultural Product Theft 937
Jin Su Kim, Min-Gu Kim, Byung Rae Cha and Sung Bum Pan

A Study of Simple Classification of Malware Based on the Dynamic API Call Counts 944
Jihun Kim, Seungwon Lee, Jonghee M. Youn and Haechul Choi

A Low-Power Sensing Management Method for Sustainable Context-Awareness in Exclusive Contexts 950
Dusan Baek, Jae-Hyeon Park, Byungjeong Lee and Jung-Won Lee

Content-Based Conformance Assurance Between Software Research Documentation and Design Guideline 957
Jong-Hwan Shin, Du-San Baek, Byungjeong Lee and Jung-Won Lee

Development of the Vision System and Inspection Algorithms for Surface Defect on the Injection Molding Case 963
Ji Yeon Lee, Wonwoo Bong, Sangjoon Lee, Chang Ho Han and Kuk Won Ko

Implementation of the Smart System for Monitoring the PCG 971
Sunho Kim, Kangwoo Lee and Yonghee Lee

The Effect of Introducing Small Cells in Wireless Networks 976
Soohyun Cho

An Enhanced Reliable Message Transmission System Based on MQTT Protocol in IoT Environment 982
Hyun Cheon Hwang, Ji Su Park, Byeong Rae Lee and Jin Gon Shon

Implementation of a Smart IoT Factory Using an Agricultural Grade Sorting Device 988
Seokhoon Jeong, Ji Yeon Lee, Kuk Won Ko and Sangjoon Lee

Segmentation and Counting of Cell in Fluorescence Microscopy Images Using Improved Chain Code Algorithm 997
Yeji Na, Sangjoon Lee, Jonggab Ho, Hwayung Jung, Changwon Wang and Se Dong Min

Empirical Study of the IoT-Learning for Obese Patients that Require Personal Training 1005
Seul-Ah Shin, Nam-Yong Lee and Jin-Ho Park

Detection of Optimal Activity Recognition Algorithm for Elderly Using Smartphone 1013
Changwon Wang, Sangjoon Lee, Jonggab Ho, Yeji Na and Se Dong Min

Method of Detecting Malware Through Analysis of Opcodes Frequency with Machine Learning Technique 1019
Sang-Uk Woo, Dong-Hee Kim and Tai-Myoung Chung

Study of Big Data Analysis Procedures 1025
Joon Ho Park, Jin Ho Park and Nam Young Lee

Design and Implementation of Authentication Information Synchronization System for Providing Stability and Mobility of Wireless Authentication 1031
Yong-hwan Jung, Jang-won Choi, Hyung-ju Lee, Joon-Min Gil and Haeng-gon Lee

Study on the Generic Architecture Design of IoT Platforms 1039
Mi Kim, Nam Yong Lee and Jin Ho Park

A Study on Digitalization of Seafarer's Book Republic of Korea for e-Navigation: Focusing on Wireless Network 1046
Jun-Ho Huh

A PMIPv6-Based Auxiliary Mobility Management Considering Traffic Locality 1053
Ki-Sik Kong

A Study on Worker's Positional Management and Security Reinforcement Scheme in Smart Factory Using Industry 4.0-Based Bluetooth Beacons 1059
SangIl Park and SeoukJoo Lee

A Study of the Extended Definition of Relation for Research Content Based Traceability 1067
Jong-Won Ko, Jae-Young Choi and Young-Hwa Cho

Transforming Algorithm of 3D Model Data into G-code for 3D Printers in Distributed Systems 1074
Sungsuk Kim and Sun Ok Yang

Cache Aware Web-Based Dynamic Adaptive Streaming Algorithm in Information Centric Networks 1079
Geun-Hyung Kim

Background Subtraction Framework for Mobile 3D Sensor Data 1086
Seongjo Lee, Seoungjae Cho, Nguyen Trong Hieu, Phuong Chu and Kyungeun Cho

A Method for Multi-user Re-identification in Invoked Reality Space . . 1091
Yunji Jeong, Yulong Xi, Jisun Park, Kyhyun Um and Kyungeun Cho

A Design Scheme of Combined Syllable Fonts for Hunminjeongeum 1096
Jeongyong Byun, Seongbum Hong and Hoyoung Kim

Proposal of a Resource-Monitoring Improvement System Using Amazon Web Service API 1103
Kyu Ik Kim, Musa Ibrahim M. Ishag, Myungsic Kim, Jin Suk Kim and Keun Ho Ryu

Author Index 1109

Advances in Information Technologies and Applications

Design and Development of a Robotic Arm for Rehabilitation and Training

Sarut Panjan(✉) and Siam Charoenseang

Institute of Field Robotics, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
sa_panjan@icloud.com, siam@fibo.kmutt.ac.th

Abstract. This proposed research proposes the design and development of a robotic arm for rehabilitation and training. This wearable robot arm can easily be used with either the user's left or right arm. Each joint of exoskeleton motion can be rotated from −90° to 90° which cover all motions of human arm. The virtual reality technique is used to provide the 3D graphics to motivate the patient during the rehabilitation. The designed robot arm will support the patient's arm during rehabilitation or training which is a repetitive task over a period of time. Furthermore, this exoskeleton arm can be utilized for training purpose.

Keywords: Exoskeleton · Rehabilitation · Human-robot interface · Virtual reality

1 Introduction

Generally, stroke rehabilitation involving with motor skill recovery is a repetitive task during a period of time. This kind of repetitive task also requires the physiotherapist to assist the stroke patient. The utilization of robot arm would be fit to do repetitive task and reduces the workloads of physiotherapist or trainer. Recently, many robots have been used in rehabilitation and training tasks. In general, robot for the rehabilitation task can be categorized into two styles which are wearable and non-wearable. Joints and links of the wearable robot are usually designed correspondingly to the patient's upper limb. For the examples, the design of "CADEN" [1] applies the cable-pulley to transmit torque/force from motor to each joint. This design can reduce weight of the exoskeleton. ARMinIII [2] is designed to have 3 degrees of freedom for shoulder and 1 degree of freedom for elbow. Robotic Upper Extremity Repetitive Trainer (RUPERT) [3] has five actuated degrees of freedom for each joint which is driven by compliant pneumatic muscle actuators (PMA). The Mechatronics and Haptic Interfaces (MAHI) [4] is designed to have 5 degrees of freedom for the elbow and forearm. The wrist of this exoskeleton is a 3-RPS (revolute-prismatic-spherical) joint. The non-wearable rehabilitation robots are usually adapted from the industrial robots. Nevertheless, this type of the rehabilitation robots has only one point of physical contact between patient wrist and the robot's end-effector. Some researchers developed non-wearable rehabilitation robots such as the MIT-MANUS [5] and MIME [6] devices. Both robots are designed for rehabilitation of shoulder and elbow joints. Benefits of wearable robotic rehabilitation are controlling and generating of force feedback to each user arm's joint. Those

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_1

rehabilitation robots are usually designed for right or left arm only. On the other hand, the advantage of non-wearable robot is flexibility for rehabilitation but it cannot control or generate force feedback to all joints at the same time.

Hence, this project proposes the development of a wearable robotic arm for rehabilitation and training. The robot should be able to hold and guide the user's arm through the predefined trajectory. Further, the user can wear this exoskeleton arm and move it to interact with the virtual object in the graphics user interface which presented on a computer monitor or head-mounted display. This proposed robot arm will support the patient's arm during rehabilitation which is repetitive task and takes a period of time.

2 System Overview

Configuration of the proposed system is shown in Fig. 1. This system consists of a universal exoskeleton arm used to send the force feedback and motion during rehabilitation and training. User will receive perceived visual and force feedbacks. The user also sees three-dimensional graphics when the user works with the robot arm. Controller is responsible for receiving commands from the main computer and controlling the robot arm's movement. Main computer takes care of processing robot commands sent to the controller and generating 3-D computer graphics accordingly with robot's movements.

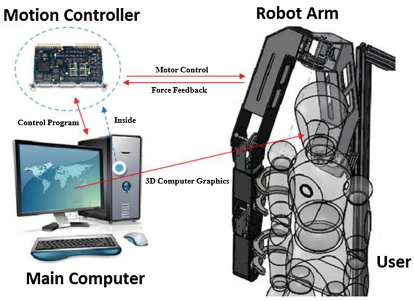


Fig. 1. System overview of robot arm control

3 Design and Implementations

The proposed robot arm is designed to have 4 degree-of-freedoms (DOF). Joint 1, 2, and 3 represent shoulder joints and joint 4 represents elbow joint. Link 1 and link 2 are aligned with the upper limb. The design will allow joint 1 and joint 2 to hold the position of shoulder while giving the shoulder some limited rotations. This will prevent the slip occurred at the shoulder. Furthermore, this robot arm is designed to be used with the left arm or right arm by rotating joint 2 about 180° as shown in Figs. 2 and 3. In the design, the cable is used to transfer the power from the servo motor to each joint of the robot arm. Therefore, this robot arm has lightweight because the servo motors are not mounted on the robot arm.

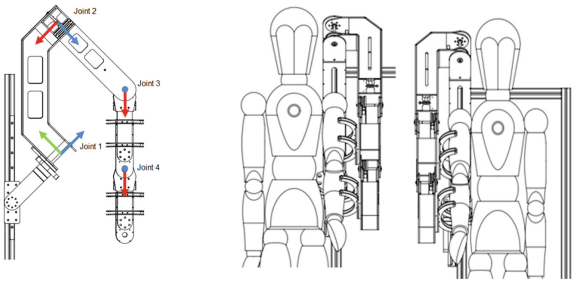


Fig. 2. Configuration of robot arm

For safety, the human arm supporters mouthed on robot arm with magnets so that they can be detached from the robot arm if any accident occurs. Furthermore, an emergency switch is included to stop all motions of robot arm when it is needed. For the brain of this exoskeleton arm system, it utilizes the main computer in order to control the designed robot arm and display 3D computer graphics feedback to the user via a computer display or a head-mounted display which is an Oculus Rift display [3]. Robot arm's joint angles and force feedback are read via NI motion control and sent to the main computer for updating the corresponding computer graphics. The UNITY 3D [4] is implemented to generate some virtual environment and display the motion of virtual arm. Main control loop is utilized by LABVIEW [5] programming and sends joint's angles to the 3D graphics manager which is based on the UNITY 3D via UDP communication. The program control of this arm system can be illustrated in Fig. 4.

The 3D graphics manager is responsible for rendering the virtual arm in the virtual environment using UNITY 3D. The joint's angles of virtual arm are updated from reading of real exoskeleton's joint's angles through main loop controller and the 3D graphics manager via the UDP communication. Position and orientation of virtual robot arm can be calculated from the robot's forward kinematics. Oculus Rift display is used for displaying all 3D graphics to the user. It can track the user's head orientation and provide 3D stereoscopic image illustrated in Fig. 5 so that the user can get some depth information, too.

For safety purpose, all experiments in this research were conducted with the human model instead of real human. In this exoskeleton system, there are three types of

Fig. 3. Configuration of robot used for left or right arm

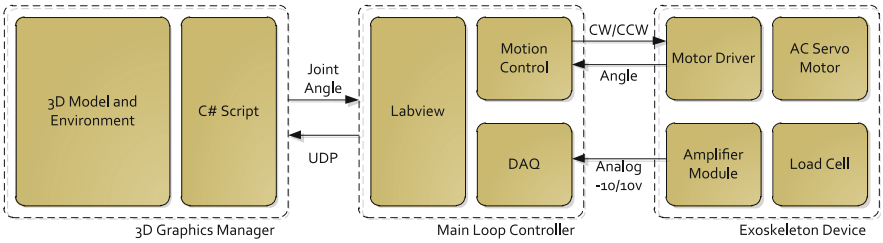


Fig. 4. Exoskeleton program control diagram

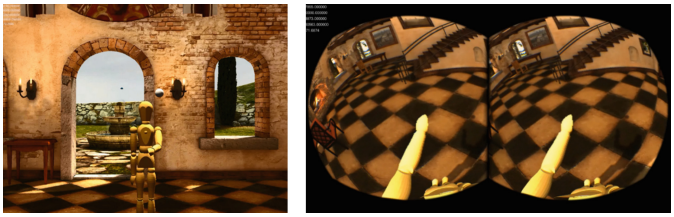


Fig. 5. Virtual environment display on oculus rift

gestures to be trained and tested. First gesture is to move the user's arm, human model's arm in this case, from right to left for the right arm configuration or from left to right for the left arm configuration. Second gesture is to move the model's right or left arms from the lower to upper position. Third gesture is to move right arm from the lower right position to the upper left position or move the left arm from the lower left position to the upper right position.

4 Experimental Results

After the study of the statistical distribution of human arm joint angles [6], the proposed exoskeleton arm was then designed to have 4 degrees of freedom which are 3 DOFs at shoulder and 1 DOF at elbow. Each joint of exoskeleton motion can be rotated from −90° to 90° which can cover all motions of human arm as shown in Figs. 7, 8 and 9. After three sets of experiments were conducted, the joint's angle and velocity trajectories of three gestures can be plotted. The right arm's trajectories are shown in Figs. 7, 8 and 9. All trajectories were predefined and replayed to control the human model to move accordingly to those predefined paths. The controller can take care of arm's load and guide the arm from the starting point to the destination for both arms successfully.

Each joint's velocity of exoskeleton arm is about 60 rpm. Moreover, there levels of safety were implemented. For mechanical safety, the arm was designed to be moved within the range of −90° to 90°. For electrical safety, the large emergency switch was added to stop all servo motors' motions if it is pressed. For software safety, limited joint's angle command was implemented. Furthermore, the admittance control and

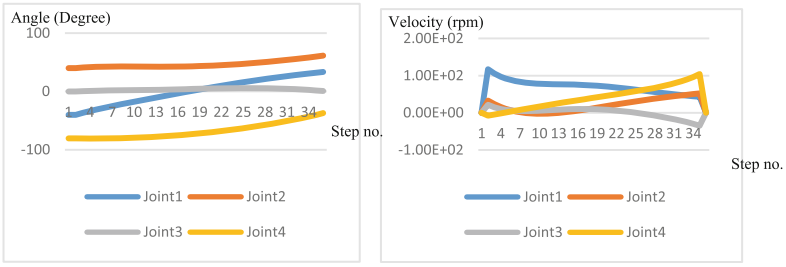


Fig. 7. Gesture 1: joint's angle trajectory (left) and joint's velocity trajectory (right)

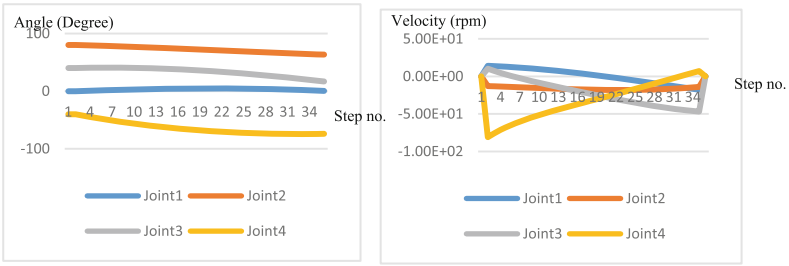


Fig. 8. Gesture 2: joint's angle trajectory (left) and joint's velocity trajectory (right)

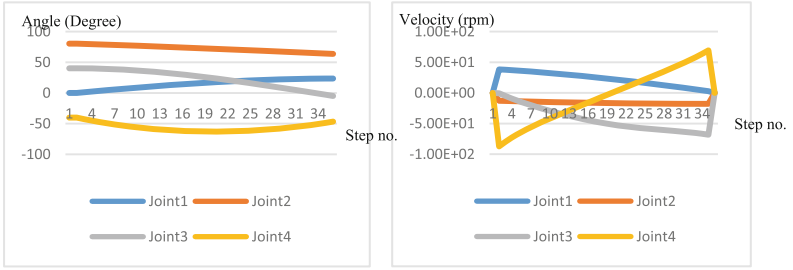


Fig. 9. Gesture 3: joint's angle trajectory (left) and joint's velocity trajectory (right)

gravity compensation were implemented so that the robot arm could be controlled to move accordingly with the forces exerting from the user. However, this active mode were not conducted at this time due to safety issue.

5 Conclusions

The 4-DOF robotic exoskeleton arm was designed and built at the Institute of Field Robotics, King Mongkut's University of Technology Thonburi. This exoskeleton arm is used as an assistive device which reduce the workload of the physiotherapist during rehabilitation of the stroke patient for recovering the motor skill. Therefore, this robot

arm can rotate cover the normal range of human's joint motion. The arm can be configured by rotating joint 2 about 180° for the use of right or left arm. This proposed system has two control modes which are active and passive modes. In this current implementation, the passive mode was conducted to move the human model to anywhere in the robot's workspace. For safety issue, the active mode was not conducted since it requires the real human to wear and move the exoskeleton arm around in the workspace. However, all force control and gravity compensation are ready for this system. In addition, the virtual environment rendered by the Unity engine was generated and integrated with this system. It can display computer graphics of virtual arm and objects through the computer display and 3D stereoscopic image through the Oculus Rift display. The 3D graphics manager can generate different scenes for any specific task.

References

1. Perry, J.C., Rosen, J., Burns, S.: Upper-limb powered exoskeleton design. IEEE/ASME Trans. Mechatron. **12**, 408 (2007)
2. Nef, T., Guidali, M., Riener, R.: ARMin III–arm therapy exoskeleton with an ergonomic shoulder actuation. Appl. Bionics Biomech. **6**, 127–142 (2009)
3. Balasubramanian, S., Wei, R., Perez, M., Shepard, B., Koeneman, J., Koeneman, E., He, J.: RUPERT: an exoskeleton robot for assisting rehabilitation of arm functions. In: Virtual Rehabilitation, pp. 163–167. IEEE (2008)
4. Gupta, A., O'Malley, M.K.: Design of a haptic arm exoskeleton for training and rehabilitation. IEEE/ASME Trans. Mechatron. **11**, 280–289 (2006)
5. Hogan, N., Krebs, H.I., Charnnarong, J., Srikrishna, P., Sharon, A.: MIT-MANUS: a workstation for manual therapy and training. I. In: Proceedings of the IEEE International Workshop on Robot and Human Communication, pp. 161–165. IEEE (1992)
6. Lum, P.S., Burgar, C.G., Van der Loos, M., Shor, P.C.: MIME robotic device for upper-limb neurorehabilitation in subacute stroke subjects: a follow-up study. J. Rehabil. Res. Dev. **43**, 631 (2006)

Detection of a Robust High-Frequency Range via Noise Analysis in a Real-World Environment

Myoungbeom Chung[1(✉)] and Ilju Ko[2]

[1] Division of Computer Engineering, Sungkyul University, Anyang City, Gyeonggi 14097, South Korea
nzin@sungkyul.ac.kr
[2] Department of Global Media, Soongsil University, 369 Sangdo-Ro, Seoul 06978, South Korea
andy@ssu.ac.kr

Abstract. Recently, many studies have been conducted on using inaudible high frequencies for wireless communication based on smart devices and data transmission algorithms. However, many studies have identified a problem such that transmission accuracy is extremely low because of ambient noise in the real-life environment. To solve this problem, we proposed an application and server system. The proposed application can gather many sounds, including those with high frequencies; the gathered high frequencies are sent to a server system that can detect a robust high-frequency range via statistical processing. We tested the proposed application's ability to gather noise and high frequencies for a certain period of time to evaluate performance. According to the testing results, the proposed application and server system could detect a robust high-frequency range via noise analysis in real life. Therefore, the proposed application and server could be a useful technology for future research on inaudible high frequencies.

Keywords: Smart device · High-frequency communication · Robust high frequency · Wireless communication

1 Introduction

Recently, because of advancements in wireless communication technology and smart device hardware, many studies have been carried out on data-correcting technologies based on smart devices. Moreover, related research has evaluated the user work rate of these technologies using smart devices and smart bands [1, 2]. In addition, many studies on user movement and activity analysis related to technologies have used various built-in sensors for smart devices [3, 4]. Smart device usage analysis focusing on user behavior has been carried out as applications installed in smart devices [5]. Most of these studies have made use of server coupling technology for big data analysis [6]. For example, user movement and activity analysis technologies send correcting data about many people's movements and location to a server. And then, analysis of information on a server can be employed for various service models, such as hotspots,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_2

favorite routes, and so on [7]. In addition, smart device usage analysis has been carried out, such as when the user sends usage data about the smart devices to the server and the server makes a decision as to whether the user is a smartphone addict by analyzing the receiving data [8].

In another stream of research on smart devices, many new communication and transmission technologies have been proposed using the devices' built-in speakers and microphone. Kim proposed user certification technology between personal computers and smart devices using high frequencies in the audible frequency range of 18–22 kHz [9]. In addition, Bihler developed a smart museum guide application using piezo speaker and smart device [10]. Chung proposed not only smart advertising service technology at a short distance, but also multi–data sharing technology among several smart devices using high frequencies [11–13]. However, because these technologies only use high frequencies that are inaudible to people, these frequencies often receive interference from the ambient that is present real life. Thus, the data transmission accuracy of Bihler's application was only 50 % when used in a real-world setting; moreover, Chung's proposed method would not work well in an actual environment because the test environment only comprised fixed noise from a speaker.

In this paper, we propose an application and server system for the detection of a robust high-frequency range via noise analysis in the real-life environment. The application uses a built-in microphone to gather ambient noise and analyze high frequencies in it. To gather and analyze high frequencies using the built-in speakers of smart devices, we developed the application based on a fast Fourier transform (FFT) algorithm. When a user carries some day with a smart device, the application can gather all noise and analyze high frequencies around the user anywhere and at any time. The proposed server system receives the information on high frequencies gathered by smart device and detects a robust high-frequency range through statistical processing after saving the high frequencies to a database. To confirm performance of the proposed application and server system, we carried out an experiment over the course of seven days. According to the result of the experiment, we can state that the proposed application and server system was able to detect the robust high-frequency range without interference from ambient noise in the real-world environment. Thus, information on the robust high-frequency range from the proposed application and server system can be used to increase the data-transmission accuracy in many studies using high frequencies.

The present paper is organized as follows: In Sect. 2, we describe the scene organization and workflow of the proposed application and server system for the detection of the robust high-frequency range. In Sect. 3, we explain the experiment to confirm the performance of the proposed system; a conclusion and a discussion of future research are provided in Sect. 4.

2 An Application for the Analysis of Ambient Noise and a High-Frequency Range Detection System

In this section, we explain the proposed system, which analyzes high frequencies from ambient noise using a smart device and sends the high frequency to the server for detection of a robust high-frequency range. Figure 1 shows the workflow of the proposed application and server system.

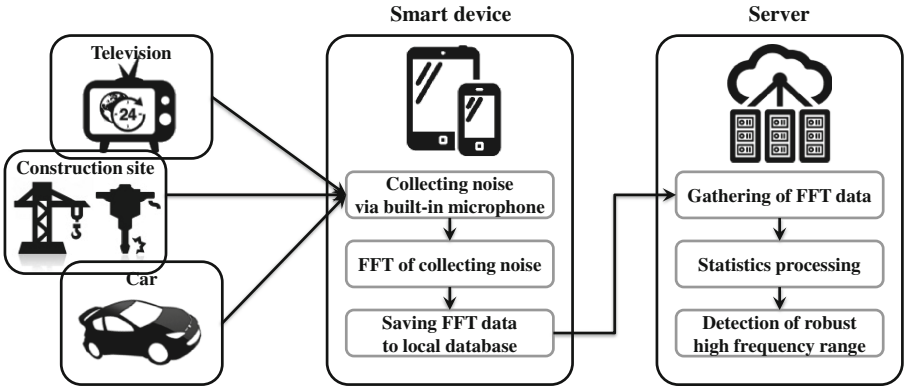


Fig. 1. The workflow of the proposed system

In Fig. 1, the application of the smart device collects various types of noise, such as television sounds, construction noise, vehicle sounds, and so on in a real-world environment via the smart device's built-in microphones. The application processes the gathered noise to generate FFT data in real time, and it selects high-frequency data from frequencies of 15–22 kHz, which are inaudible to people. Following this, the application saves the selected high-frequency data to the smart device's local database; the high-frequency data are saved up to a fixed quantity, at which point the application sends the data to the server system and deletes them. Next, the server system collects high-frequency data from smart devices and carried out basic statistical analysis. Thus, the proposed system detects a robust high-frequency range in real time through this process.

The proposed application gathers high-frequency data by 100 Hz; it saves high-frequency data to the database after confirming that high frequencies are present according to number of frequency bins. The reason for division by 100 Hz is that most of the existing research has used such division, and it can avoid interference from each high frequency. Thus, the application gathers high-frequency data from ambient noise and carries out FFT processes. Next, the server system saves received high-frequency data from the smart device to a data collection table with the collection time. Simultaneously, the server saves the result concerning whether high frequencies are present for each datum to the statistics table. When the server saves the high-frequency data to the data collection table, it does not process any data, and it saves the original number of high frequencies from 15–22 kHz from the smart device. The data collection table has

Table 1. Data collection table schema for saving FFT data received from the smart device.

Schema	Type	Description
no	int(10)	Number of counting of FFT data
time	int(13)	Time stamp when FFT data were saved on smart device
s150	int(2)	Frequency bin value of FFT data at 15.0–22.0 kHz
… …	… …	
s220	int(2)	
reg_date	int(13)	Time stamp when FFT data from smart device was saved

the schema shown in Table 1. In Table 1, "no" counts the row number of FFT data, and "time" means the saving time of the FFT data on the smart device. Moreover, s150 is the number of high-frequency bins at 15.0 kHz, and s220 is the number of high-frequency bins at 22.0 kHz. Finally, reg_date refers to the time FFT data were saved in this server table. The statistics table for the detection of the robust high-frequency range has the same schema as in Table 1. However, this table does not have a time field, which means that the time stamp for the FFT data is saved on the smart device.

Because the statistics table only counts whether there is a high frequency, when the server saves high-frequency data to the data collection table, the statistics table is updated by counting the existing high-frequency data. Then, when we count the number related to each high frequency from the statistics, we were able to identify some high frequencies that often occur in the real-world environment; moreover, we could detect the robust high-frequency range from rare high frequencies in the statistics table.

3 Experiments and Evaluation Using the Proposed Application and Server System

This section introduces an application developed to gather ambient noise around smart devices; here, we analyze the experiments and evaluate the proposed application and server system. First, we developed the application based on iOS; Fig. 2 shows the main screen of the application. In Fig. 2, the graph located on the left is the bin number of each high frequency from the ambient noise gathered by the built-in microphone; we can confirm high frequencies at 16.7, 16.8, 18.1, 18.2, 18.9, 19.0, 19.1, 19.2, and 20.7 kHz. The application analyzes current ambient noise currently and shows the graph according to the FFT data when the user touches the "Frequency check" button located on the right. To analyze the FFT data, for this experiment, we used a 48,000 sampling rate and the FFT library provided by Baoshe Zhang [14]. When a user touches the "Start gathering noise" button, the application starts collecting and analyzing ambient noise at regular intervals, and the button is changed to "Stop gathering noise." The total duration of ambient noise collection is shown as the GTime, and the total duration of the main screen is 28 min and 21 s. When the user touches "Send gathering data" button, the application sends the total FFT data collected up to that time to the server system; it shows the final sending time at "Last send" located at the bottom right. Moreover, when the volume of collected data exceeds a specified amount,

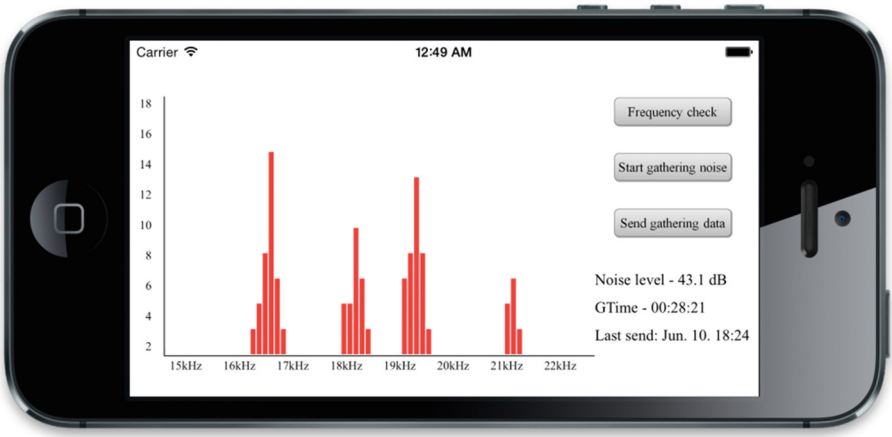


Fig. 2. Main screen of the proposed application for collecting of ambient noise

the application sends the FFT data to the server automatically even if the user has not touched the "Send gathering data" button.

We carried out an experiment on the detection of the robust high-frequency range using the proposed application and server system. A participant used a smart device running the proposed application in everyday life for 7 d. Running time of the application was from 9:00 to 10:00, 12:00 to 13:00, and 18:00 to 19:00; thus, the total running time over the 7 d was 21 h. The participant's working area varied during the experiment, including home, cafés, a laboratory, the subway, bus stop stations, crowded buses, and so on. The application collected and analyzed ambient noise at 2-min intervals and the participant touched "Send gathering data" button at each ending time to send the collected FFT data to the server. The server systems for the FFT data saving and statistics processing were Apache 2.2.14, PHP 5.2.12, and MySQL 5.1.39, and the server system comprised an Intel® Core™ i5 CPU 750 and 8G RAM. Through this experiment, the data collection table saved 630 records about high-frequency data from ambient noise in a real-world environment; Fig. 3 gives an example of some records in the data collection table. In Fig. 3, the "time" field is the saved time of high-frequency data of the row as a time stamp. For example, the time value of no 1 is 1463356872, referring to 09:01:12, May 16th, 2016.

A total of 71 fields were present in the data collection table from save frequency bins of s150 to s220 for each high frequency, and Fig. 3 shows a sample of these fields. In Fig. 3, the value of reg_date from no 1 to no 10 is the same; this is because reg_date is the time at which high-frequency data were saved from smart device to server, and rows 1–10 were saved at the same time. Below, Fig. 4 shows the processing result of statistics table, which determines whether each high frequency is present from the data collection table.

Figure 4 represents a statistical graph concerning high frequencies that the participant gathered over 21 h; we can see that some high frequencies were not evident, such as 15.0–15.3, 15.9–16.2, 16.4–16.6, 17.3–17.6, 18.4–18.6, 18.8, 19.0–19.2, 19.4,

no	time	s150	s155	s160	s165	s170	s175	s180	s185	s190	s195	s200	s205	s210	s215	s220	reg_date
1	1463356872	1	1	1	0	3	2	0	0	0	1	0	0	0	4	0	1463360475
2	1463356992	0	3	0	4	6	0	0	0	0	1	0	0	0	0	0	1463360475
3	1463357112	0	3	0	4	6	0	0	0	0	1	0	0	0	0	0	1463360475
4	1463357232	0	0	0	0	3	0	6	0	0	3	3	0	0	0	0	1463360475
5	1463357352	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1463360475
6	1463357472	0	0	0	0	0	0	0	0	0	0	11	0	0	4	0	1463360475
7	1463357592	0	11	0	4	0	0	0	0	0	0	0	0	0	0	0	1463360475
8	1463357712	0	0	0	2	0	0	2	0	4	0	0	0	0	0	0	1463360475
9	1463357832	0	3	0	4	6	0	0	0	0	1	0	0	0	0	0	1463360475
10	1463357952	0	0	0	0	0	0	12	0	0	1	0	0	2	1	0	1463360475

Fig. 3. Example of FFT data collection from a smart device

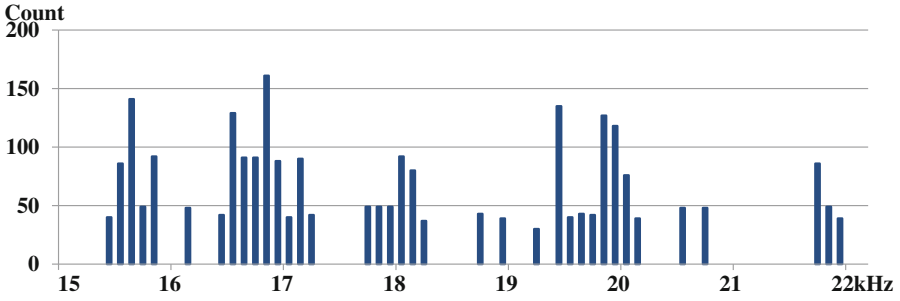


Fig. 4. The high frequency chart from the statistics table of the server

20.3–20.5, 20.7, and 20.9–21.6 kHz. Thus, we could detect robust high-frequency ranges that hardly ever occur from ambient noise when we use the proposed application and server system. Moreover, we consider that the detected robust high-frequency range could not receive interference from ambient noise, which would occur in real life when many researchers sought to use the range.

4 Conclusions and Future Work

In this paper, we proposed an application that can gather and analyze high frequencies from ambient noise using the built-in microphone of a smart device and a server system that can save high-frequency data to database and process statistics from the collected data for the detection of the robust high-frequency range. From the experiment, we confirmed that the proposed application could analyze high frequencies from ambient noise in real time and send the FFT data to the server. In addition, we showed that the server system can distinguish high frequencies that are often present from others that hardly ever occur in the statistics table. Therefore, the proposed application and server system would be useful for the detection of the robust high-frequency range from ambient noise, and we expect that they can be used as effective technology in the data communication and transmission field based on smart devices using high frequencies.

Finally, in future research, we will study the analysis of the threshold α analyze in relation to the propriety of the value when changed, because we only used a threshold of 5 in this paper. Moreover, because we only tested the system 3 h per day over 7 d, we need to collect more data for statistical processing and carry out more testing. Therefore, we will perform many experiments with the proposed application and server system over longer periods, and we will study statistical processing based on big data.

Acknowledgments. This research project was supported in part by the Ministry of Education under Basic Science Research Program (NRF-2013R1A1A2061478) and (NRF-2016R1C1B 2007930), respectively.

References

1. Park, J.T., Hwang, H.S., Moon, I.Y.: Study of wearable smart band for a user motion recognition system. Int. J. Smart Home **8**(5), 33–44 (2014)
2. Jhajharia, S., Pal, S.K., Verma, S.: Wearable computing and its application. Int. J. Comput. Sci. Inf. Technol. **5**(4), 5700–5704 (2014)
3. Bo, C., Zhang, L., Li, X.Y., Huang, Q., Wang, Y.: Silentsense: silent user identification via touch and movement behavioral biometrics. In: 19th Annual International Conference on Mobile Computing & Networking, pp. 187–190. ACM, Miami (2013)
4. Wu, X., Brown, K.N., Sreenan, C.J.: Analysis of smartphone user mobility traces for opportunistic data collection in wireless sensor networks. Pervasive Mob. Comput. **9**(6), 881–891 (2013)
5. Falaki, H., Mahajan, R., Estrin, D.: SystemSens: a tool for monitoring usage in smartphone research deployments. In: 6th International Workshop on MobiArch, pp. 25–30. ACM, Bethesda (2011)
6. Chittaranjan, G., Blom, J., Gatica-Perez, D.: Mining large-scale smartphone data for personality studies. Pers. Ubiquit. Comput. **17**(3), 433–450 (2013)
7. Abe, M., Fujioka, D., Handa, H.: A life log collecting system supported by smartphone to model higher-level human behaviors. In: 6th International Conference on Complex, Intelligent and Software Intensive Systems, pp. 665–670. IEEE, Palermo (2012)
8. Mok, J.Y., Choi, S.W., Kim, D.J., Choi, J.S., Lee, J., Ahn, H., Song, W.Y.: Latent class analysis on internet and smartphone addiction in college students. Neuropsychiatric Dis. Treat. **10**, 817–828 (2014)
9. Kim, J.B., Song, J.E., Lee, M.K.: Authentication of a smart phone user using audio frequency analysis. J. Korea Inst. Inf. Secur. Cryptology **22**(2), 327–336 (2012)
10. Bihler, P., Imhoff, P., Cremers, A.B.: SmartGuide: a smartphone museum guide with ultrasound control. Procedia Comput. Sci. **5**, 586–592 (2011)
11. Chung, M.B., Choo, H.S.: Near wireless-control technology between smart devices using inaudible high-frequencies. Multimedia Tools Appl. **74**(15), 5955–5971 (2015)
12. Chung, M.B.: An advertisement method using inaudible sound of speaker. J. Korea Soc. Comput. Inf. **20**(8), 7–13 (2015)
13. Chung, M.B.: Effective near advertisement transmission method for smart-devices using inaudible high-frequencies. Multimedia Tools Appl. **75**(10), 5871–5886 (2016)
14. Baoshe, Z.: Java FFTPack. http://jfftpack.sourceforge.net

Intelligent Food Distribution Monitoring System

Ganjar Alfian[1(✉)], Hyejung Ahn[2], Yoonmo Shin[1], Jaeho Lee[1], and Jongtae Rhee[2]

[1] u-SCM Research Center, Nano Information Technology Academy, Dongguk University-Seoul, Seoul 100-272, Korea
{ganjar, ymshin, rapidme}@dgu.edu
[2] Department of Industrial and Systems Engineering, Dongguk University-Seoul, 26, 3-ga, Pil-dong, Chung-gu, Seoul, Korea
{macarori, jtrhee}@dgu.edu

Abstract. Food quality and safety has gained main attention, due to increasing health awareness of customer, improved economic standards and lifestyle of modern societies. Thus, it is important for consumers to purchase good quality products in order to keep the customer satisfaction level. In this study, we propose traceability system for food by monitoring the location as well as temperature and humidity. The RFID technology and wireless sensor network are utilized in this study to perform the experiment. The real testbed implementation has been performed in one of the Korean Kimchi Supply Chain. The result showed that our proposed system gave the benefit to the manager as well as customer by providing real time location as well as temperature-humidity history. It will help manager to optimize the food distribution while for the customer it will increase the satisfaction by maintaining the freshness of product.

Keywords: Monitoring system · RFID · Temperature and humidity sensor

1 Introduction

Food quality and safety has gained main attention, due to increasing health awareness of customer, improved economic standards and lifestyle of modern societies. Thus, it is important for consumers to purchase good quality products in order to keep the customer satisfaction. In South Korea, there have been a number of cases reported by Korea Agro-Fisheries & Food Trade Corporation related to illegal agricultural food distribution channels that disguised cheap imported agricultural food products as the quality products into the local food chains [1]. In addition, the economic cost of foodborne illness in the USA alone is 50 billion to 80 billion dollars annually; it includes the health care costs, lost productivity, and diminished quality of life [2]. Therefore, in agriculture food industry, the strict control and monitoring of food freshness and quality is very important.

The RFID (Radio Frequency Identifier) technologies promise to revolutionize future inventory management [3]. The RFID applications hold tremendous promise;

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_3

such as RFID tag monitoring. Tag monitoring is a key component in applications like item tracking and inventory control. By frequently scanning its inventory, a retailer can quickly determine if anything is missing, and act accordingly. In addition, the RFID technology has been implemented successfully in Supply Chain Management to reduce the inventory losses [4], for identification and monitoring of agricultural animals [5, 6], food traceability system [7], and healthcare applications to improve patient care as well as reduce overall costs [8].

Furthermore, Wireless sensor networks (WSN) are spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to cooperatively pass their data through the network to a main location. By utilizing the sensor, the history of room temperature and humidity can be presented in the system as real time [9].

In addition, to monitor the freshness of product, the time temperature indicator (TTI) can be used also to monitor the freshness of product. A TTI is a device or smart label that shows the accumulated time-temperature history of a product. Time temperature indicators are commonly used on food, pharmaceutical, and medical products to indicate exposure to excessive temperature. The application of TTI for food safety management has been successfully demonstrated in previous study [10].

This study proposes *E-Pedigree Food Tracking System* which is integrated with EPCIS (Electronic Product Code Information System). *E-pedigree* or electronic pedigree is an electronic document which provides data on the history of a particular batch of a drug. In this case, we use this pedigree system to protect the quality of food product in order to avoid the illegal agriculture food distribution channels. EPCIS is an EPC global standard for sharing EPC related information between trading partners. EPCIS provides standard for capturing and communicating the business events for tracking and tracing products within an enterprise and across the supply chain. In addition, the *Real-time Temperature-Humidity Sensor* also will be developed, thus the temperature and humidity of product can be monitored. The *TTI QR Code based Scanner* will be developed as addition, to recognize the color of TTI, thus the product freshness can be presented to customer. By integrating these three systems, it is expected to assure the quality and safety of agriculture food products throughout the whole supply chain, so that the customer can check the quality of product whenever he/she buys the agriculture food product.

2 Methodology

In this study, we are considering Kimchi distribution network which consists from Kimchi Producer, Transporter, Distributor, and Customer. The pedigree system for agriculture food products (i.e., kimchi) based on the EPC global network is developed. With the usage of RFID, EPC global architecture and wireless sensor networks (WSN), we can not only track and trace the product across the complete supply chain but also we can monitor the environmental information such as temperature and humidity of agriculture food product.

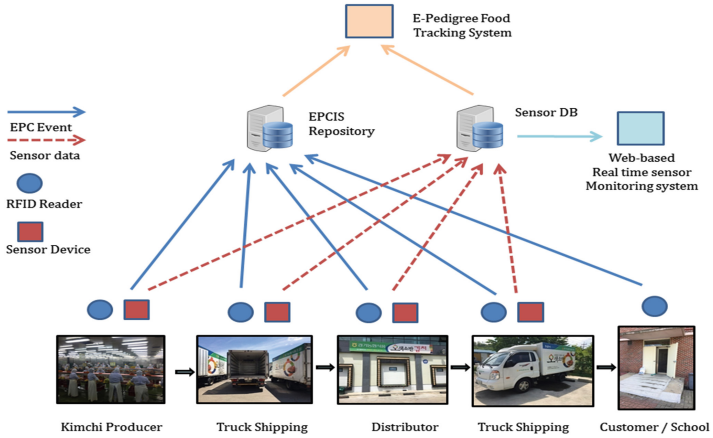


Fig. 1. Pedigree system and sensor architecture

First, for the *E-Pedigree Food Tracking System,* the history of each product is stored in the EPCIS repository. The design of the e-pedigree system can be seen as follow.

As can be seen in Fig. 1, the system architecture shows the *EPC event* and the *sensor data* flow between different supply chain partners. The supply chain partners capture the EPC code (product information) using RFID Reader and store it in EPCIS repository while the temperature-humidity data are captured by sensor device and store it into sensor database. In the end, by combining two databases from EPCIS repository and sensor database, it will create complete pedigree of product (i.e., location, process business, temperature, humidity, etc.).

In order to capture the EPC data, the RFID reader are utilized in this study. The RFID reader, antenna, antenna holder and laptop/computer are needed to gather the Tag information. The *capturing application* will be developed in this study, based on Java Programming Language to receive the data from Reader and sent it to the EPCIS repository.

Second, for *the Real-time Temperature-Humidity Sensor*, the sensor device will be installed in each room of the Kimchi Supply Chain, such as cold storage of producer, cold storage of truck (transporter), and cold storage of distributor. The real time sensor device will be developed based on android application. The sensor device is attached to the android Smartphone, the data (temperature-humidity) are handled by the android app and send it to the sensor database (in server) frequently, such as for every 5 s/1 min. Thus the changing temperature and humidity inside the room will be monitored real time and can be seen in the *web-based sensor monitoring system*.

Third, for *TTI QR Code based Scanner App*, the QR code is printed on the TTI (Time Temperature Indicator) and attached on the product (kimchi package/box), thus the changing color in TTI will reflect the color QR code as well. Traditional QR code only represented as black and white, while in this study, the QR code color can be represented as a variety of *Color Lab Value* (can be seen in Table 1). Thus the functionality of *TTI QR Code based Scanner App* we developed in this study is

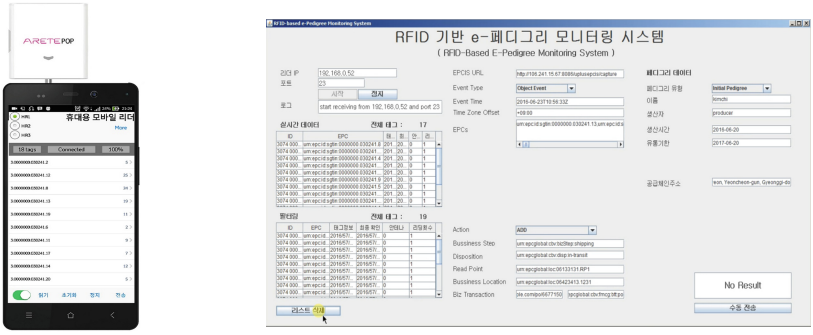
expected not only understanding hidden information of QR code but also understanding its color (color of TTI). Each information of color in QR code represents different quality of product. The detail information can be seen in Table 1. The TTI QR code which is attached on the product (such as kimchi box) will be captured by our *TTI QR Code based Scanner App*, detects the certain color and return the detail information which can reveal the product quality.

Table 1. Each color lab value of QR code represents different food quality.

TTI image	Food status	Detail	TTI description
	Good quality of product	색체계 L a b Color Lab Value 85.36333 -11.2733 -3.40667 색체계 R g b Color Rgb value 210.02 225.63 222.987	Just activated
	Still Fresh for current use	색체계 L a b Color Lab Value 84.58 -15.9433 -6.00333 색체계 R g b Color Rgb value 169.692 220.437 221.767	Still usable for food
	Almost expired	색체계 L a b Color Lab Value 82.39 -20.56 -10.67 색체계 R g b Color Rgb value 145.051 216.96 224.167	Almost expired
	Already expired	색체계 L a b Color Lab Value 79.83333 -23.1633 -14.41 색체계 R g b Color Rgb value 123.697 211.381 223.819	Already expired

3 Result and Discussion

For the implementation of *E-Pedigree Food Tracking system,* the RFID tags are attached on the Kimchi box which are delivered to the whole supply chain. The RFID reader (*ALR-9900* RFID Reader) read the tags, while our *RFID capturing application* receives the data and sends it to the server through HTTP protocol. The *RFID Capturing application* has received the tags data and presented in the left side of the program. The EPC document and additional pedigree data are generated in the center and right side of the program respectively, detail can be seen in Fig. 2(b).



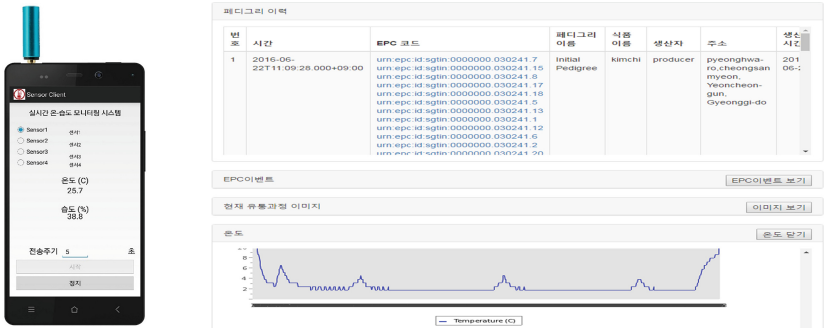
(a) (b)

Fig. 2. (a) The hand held reader and (b) RFID capturing application

In addition, the RFID reader (*ALR-9900* RFID Reader) is not suitable for flexible used, such as for truck driver. Thus in this study, we have developed *hand held RFID reader* based on android application. The RFID dongle reader (AretePop RFID Reader) is utilized to read the data, while the android app is developed to handle the data before sending it to the EPCIS server. The detail can be seen in Fig. 2(a).

For the *Real-time Temperature-Humidity Sensor*, the sensor device is used for the study. The sensor device (FTLab smart sensor) is attached to the Smartphone through the audio jack. The sensor data (temperature and humidity) is then gathered and presented on the Smartphone screen by the app we have developed. The detail can be seen in Fig. 3(a). The app can be used for different sensor functionality, i.e. in this study we utilize four sensors and installed in different location. In addition, the temperature and humidity is sent by the sensor device to the server frequently, thus the real-time temperature-humidity data of each location in Supply chain can be presented.

As can be seen Fig. 3(b), the *e-Pedigree food tracking system* based on the web is developed. The *e-Pedigree Food tracking system* present the history of product, the location and the temperature-humidity regarding the product, thus the manager can optimize the distribution of products as well as monitoring product quality.



(a) (b)

Fig. 3. (a) The handheld reader and (b) e-Pedigree food tracking system

Furthermore the *TTI QR Code based Scanner App* is developed in this study based on the android environment. Once customer open the app and put the camera of Smartphone above the TTI QR code in the correct position and good light condition, the app will automatically detect the information inside the QR code and its color. The app opens the other page which shows the detail quality of product. As explained before, the app analyzes the majority of color which contain in TTI QR code, then show the similar color which represent the quality of product. As can be seen in Fig. 4(a) and (b), the blue color of TTI QR code represents the good quality of product, safe to be consumed by customer.

(a)

(b)

Fig. 4. (a) The app scanned the TTI QR code based and (b) shows the product quality based on its color

4 Conclusion

This study has successfully implemented the *e-pedigree food tracking system*, which helped customers in order to assure the quality and safety of agriculture food products throughout the whole supply chain. By using this *e-pedigree food tracking system*, the customer can see the detail of food history from producer until retailer, thus they can believe that the certain product (or food) does not contains from illegal channels.

In addition, by utilizing wireless sensor networks (WSN), *the real-time temperature-humidity system* was developed to monitor the temperature-humidity of cold storage for whole supply chain. Thus the manager can monitor the temperature and humidity real time, and act accordingly if something happen (i.e., the temperature is not stable). In addition, by presenting the temperature-history in product pedigree, it will help the customer to understand the product quality in detail.

Furthermore, the *TTI QR Code based Scanner App* was developed to understand the color of TTI QR code which is attached on the product. The app was able to detect the information of QR code and its color which represent the quality of product. This app will help customer to analyze the color of TTI QR code and give the information of product quality to the customer easily.

Acknowledgement. This work was supported in part by the Ministry for Food, Agriculture, Forestry and Fisheries through the Agriculture Research Center under Grant 710003-03-1-SB110 and in part by the 2014 Research Fund through Dongguk University.

References

1. Korea Agro-Fisheries & Food Trade Corporation. http://www.mfds.go.kr/. Accessed 04 January 2016
2. Scharff, R.L.: Economic burden from health losses due to foodborne illness in the united states. J. Food Prot. **75**, 123–131 (2012)
3. Luo, W., Chen, S., Li, T., Chen, S.: Efficient missing tag detection in RFID systems. In: 2011 Proceedings IEEE INFOCOM. IEEE (2011)

4. Sarac, A., Absi, N., Dauzère-Pérès, S.: A literature review on the impact of RFID technologies on supply chain management. Int. J. Prod. Econ. **128**(1), 77–95 (2010)
5. Voulodimos, A.S., Patrikakis, C.Z., Sideridis, A.B., Ntafis, V.A., Xylouri, E.M.: A complete farm management system based on animal identification using RFID technology. Comput. Electr. Agric. **70**(2), 380–388 (2010)
6. Feng, J., Fu, Z., Wang, Z., Xu, M., Zhang, X.: Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. Food Control **31**(2), 314–325 (2013)
7. Hong, I.-H., Dang, J.-F., Tsai, Y.-H., Liu, C.-S., Lee, W.-T., Wang, M.-L., Chen, P.-C.: An RFID application in the food supply chain: a case study of convenience stores in Taiwan. J. Food Eng. **106**(2), 119–126 (2011)
8. Gaynor, M., Waterman, J.: Design framework for sensors and RFID tags with healthcare applications. Health Policy Technol., Available online 18 July 2016
9. Rashid, B., Rehmani, M.H.: Applications of wireless sensor networks for urban areas: a survey. J. Netw. Comput. Appl. **60**, 192–219 (2016)
10. Koutsoumanis, K.P., Gougouli, M.: Use of time temperature integrators in food safety management. Trends Food Sci. Technol. **43**(2), 236–244 (2015)

Design of Sudden Unintended Acceleration Check System Using Distance Measurement Sensor

Jea-Hui Cha, Tae-Hyoung Kim, and Jong-Wook Jang(✉)

Department of Computer Engineering, Dong-eui University,
Gaya 1-dong, Busanjin-gu, Busan, Korea
ckwpgml5507@naver.com, fingersnoop@gmail.com,
jwjang@deu.ac.kr

Abstract. Today, various models of automobiles based on diverse motifs, such as eco-friendly car, autonomous car, connected car, and smart car, are produced in Korea. With the increased competitiveness of the automobiles, highly-advanced automobiles are developed.

Although the automobiles have positive aspects, their sudden unintended acceleration (SUA) accidents bring about people's negative perception of the vehicles.

It has been analyzed that the SUA is attributable to defects of electric devices or brakes and other factors in the development of automobile parts and technology. Nevertheless, it is impossible to define the causes of the SUA accurately, and manufacturers also try to avoid their responsibility. As a result, more burden has been imposed on drivers.

Therefore, this study designs the system that has a camera and a distance sensor attached to a driver's seat and helps to respond to the SUA and find its cause in the way of checking the operation states of a vehicle's control parts including accelerator and brake and sensor images.

Keywords: OBD · Sensor · Connected car · Car · Embedded

1 Introduction

Today, automobiles with various functions are produced in Korea. Eco-friendly car, autonomous car, connected car, smart car, and other cars for diverse purposes are produced. In the circumstance, the connection of IT and automobiles is not selective, but essential. Despite the positive aspects of such a connection, malfunction caused by use complexity and electromagnetic defects may lead to a serious problem. The biggest problem of malfunction is an accident made by malfunction of a brake. In other words, the accidents made by a failure of brake in high-speed driving or by fast driving without a driver's stepping on accelerator continue to occur and are unable to be ignored. Therefore, this study focuses on the accidents caused by sudden unintended acceleration (SUA).

A SUA accident is generally attributable to ECU interference of electromagnetic field which leads to unexpected acceleration. Regarding the statistics of the SUA

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_4

accidents of each automobile brand, from 2010 to July 2015, there had been 482 accidents made by suspect SUA in Korea. The automobile company with the largest SUA accidents was Hyundai Motors, followed by Kia Motors and Renault Samsung. What is more serious is that the vehicles without airbag deployment at the time of the SUA numbered 209, accounting for 43.4 %. Therefore, in an unexpected situation, a driver's safety is not guaranteed [1].

Moreover, there is no proper compensation for SUA accidents. In Korea, most drivers who experienced the SUA came to lose a suit. That is because such an accident is judged to be a driver's mistake.

Therefore, to judge whether a SUA accident is caused by a driver's mistake, this study designed a system as follows:

A camera and a sensor are attached to a vehicle's accelerator and brake. Images of the accelerator and brake are obtained to check their control states. The intensity of stepping on the devices is measured on the basis of the images. Both the measured values and the images are saved. OBD-II information is extracted and is compared with the formation on a vehicle's current status acquired in the above way in order to determine the cause of the SUA accident.

2 Related Work

2.1 Overview of EDR (Event Data Recorder) System

If a vehicle's airbag is deployed, or if safety belt tensioner works, accident information is saved in EDR. Also, unless airbag is not deployed and a certain level of shock occurs, crash information is saved. The data measured by various sensors to a vehicle are temporarily saved in airbag detection system. Once a crash signal is detected, the driving data before the signal detection and the crash data after the detection are recorded in EDR system. 3). The EDR data recorded when airbag works are saved permanently and are unable to be deleted. The saved EDR data are analyzed in the way of connecting the CDR(Crash data retrieval) directly to a vehicle's OBD terminal or the module in which EDR data are saved. EDR data are somewhat different depending on a manufacturer or a vehicle model. Usually, the driving information saved 5 s before an accident (pre-crash data), crash data, airbag deployment information, and various kinds of system information are recorded. The data components recorded in EDR are presented in Fig. 1 [2].

pre-crash data	crash data
▪Vehicle speed ▪Number of engine RPMs ▪Location of accelerator pedal ▪Brake switch ▪Location of transmission gear ▪Angles of steering wheel ▪State of hazard light ▪State of seat belt wearing ▪Other driving information ▪Other system information	▪ Head-on collision ▪ Read-end collision ▪ Side-on collision ▪ Speed change(x/y directions) ▪ Crash acceleration (x/y directions) ▪ Roll over angle ▪ Airbag deployment information ▪ Safety belt tensioner information ▪ Other crash information

Fig. 1. Dada element of EDR [2]

With the use of EDR, it is possible to analyze types of traffic accidents, including intentional accident, running speed and collision speed, efficient collision speed, criticality of crash, and the direction of impact force. The system can be used to analyze information on the SUA. However, in the case of a suspected SUA accident, the system is unable to check a driver's information and thus is not helpful at all.

Therefore, this study designed a hardware system along with OBD-II to analyze a SUA accident and provide necessary information to a driver.

2.2 Hardware Components of Sudden Unintended Acceleration Check System

Based on the MCU MCU(Cortex-M3), the hardware system can integrate OBD-II and the data of 3 distance measurement sensors(LK-MDS-C29), can send effective data to mobile and server to save them, and enables a driver to check the data at any times.

2.2.1 MCU(Cortex-M3)

Cortex-M3 Processor is an ARMv7-M profile processor that responds to the market of existing 8Bit Microcontroller (AVR, PIC, 8051, etc.) featuring low gate count, low interrupt latency, and low-cost. Cortex-M3 can combine the commands of 16 and 32 bits for use. With no more mode switching and 16-bit code density, it can make the performance of a 32-bit command. The processor has backward compatibility with 16-bit Thumb Instruction [3].

Figure 2 illustrates performance of Cortex-M3 series.

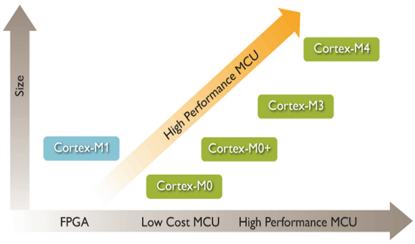


Fig. 2. Performance of Cortex-M3 series [3]

2.2.2 Distance Measurement Module (LK-DMS-C29)

Figure 3 shows the distance measurement module LK-DMS C29. The module has the distance measurement range of 10–80 cm, the infrared distance measurement sensor operation voltage of 4.5 V–5.5 V, and no need of input signals. Therefore, the device is easy to be handled, is not influenced greatly by a reflector's color and reflectivity, and makes it possible to perform high-precision measurement based on continuous distance and average operation output [4]. Because of the advantages, the module was selected.

This system used three distance measurement modules, each of which was attached to accelerator, brake, and clutch to check the intensity of each controller.

Fig. 3. Distance Measurement Module (LK-DMS-C29)

3 System Design and Main Functions

3.1 System Architecture

The integrated module sends Cortex–M3 a variety of information of OBD–II, such as SPEED, RPM, TIME, VOLT, MAF, which is generated when accelerator is stepped on. The values of the distance measurement modules connected to Cortex–M3 and camera information are integrated. The data are sent to an App via USB or Bluetooth. In Fig. 4, App can save relevant logs in Server through LTE, WIFI, and WCDMA, and check the data on web.

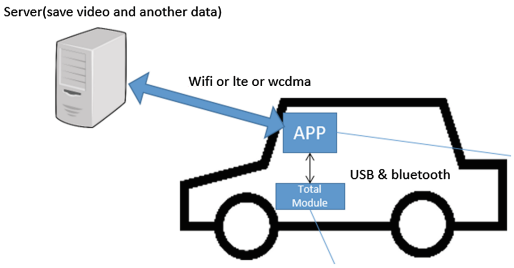


Fig. 4. System architecture

3.2 Installation of Distance Measurement Module

In this system, a distance measurement module is installed on a vehicle's accelerator with the use of ARM Cortex–M3 (Fig. 5).

The data obtained from the distance measurement module are raw data with the unit of voltage. It is necessary to measure a distance with a voltage value. The table of conversion into cm is presented below:

As shown in Fig. 6, CM values are different depending on voltage values. Therefore, it is necessary to make a mathematical formula for the table.

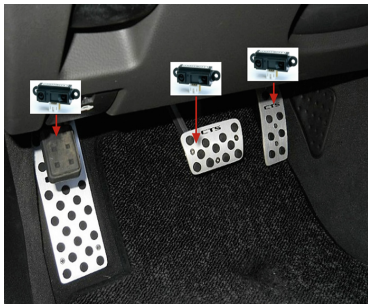


Fig. 5. Installation of distance measurement module sensor [5]

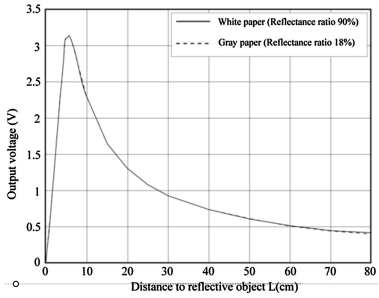


Fig. 6. Table of distance based on voltage values [4]

3.3 Camera Module

A camera module next to the accelerator photographs images in real time and saves them in Server. The image data are used as evidence of a lawsuit.

4 Conclusion

The problems of a suspected SUA accident are that it is hard to find its cause and solution and that it is very difficult to win a suit against the company of the vehicle with a SUA accident. That is because there is no evidence. It is expected that the system designed in this study is used as evidence in a lawsuit and for post-treatment and contributes to the prevention of SUA accidents.

Therefore, this study designed the system that processes a vehicle's information in real time with the use of camera, distance measurement sensor, and OBD-II, accurately finds the cause of a SUA accident, helps to provide accurate information for post-treatment, and enables a driver to check vehicle information and SUA images from server on web or an application at any times.

Acknowledgments. This Work was supported by Dong-eui University Foundation Grant (2016).

References

1. http://www.kasdi.co.kr/analysis/2830
2. Yun, D., Kim, Y., Lee, H.: Study on EDR utilization for traffic accident analysis. In: 2014 Korean Society of Automotive Engineers Conference & Exhibition, vol. 11, pp. 1439–1444 (2014)
3. http://www.jkelec.co.kr/img/lecture/cortex_arch/cortex_arch_2.html
4. http://www.lkembedded.co.kr/shop/goods/goods_view.php?goodsno=16&inflow=naver&NaPm=ct%3Diqhtejo8%7Cci%3D0ac5d96b182a1088953b5bb7e6fee6657a271f95%7Ctr%3Dsls%7Csn%3D281164%7Chk%3D288c08e9c5fd9c948fb2277c9db85d824cbae4ae
5. http://blog.naver.com/wjsdudwns125?Redirect=Log&logNo=220676025978

Real-Time Dynamic Motion Capture Using Multiple Kinects

Seongmin Baek(✉) and Myunggyu Kim

Electronics and Telecommunications Research Institute (ETRI),
218 Gajeong-ro, Yuseng-gu, Daejeon 34129, Korea
{baeksm, mgkim}@etri.re.kr

Abstract. The present paper proposes a method of capturing real-time motions without any inconvenient suit by using several inexpensive sensors vulnerable to joint occlusion and body rotation. Depth data and ICP algorithm are used for calibration. Then, the left and right sides of joints are determined, and the optimal joints are chosen based on the variation in rotation to restore postures. The similarity between the motions captured by the proposed multiple sensors and those captured by a commercial motion capture system is over 85 %.

Keywords: Multiple kinects · Motion capture · Dynamic motion · Joint selection

1 Introduction

3D motion capture has long been explored and characterized by high applicability in diverse fields for retrieving body motions. Optical systems are often used for capturing motions. Still, magnetic systems are also used to secure free movements. Yet, due to the need to wear inconvenient suits, motion capture systems are difficult to apply to ordinary users. By contrast, the marker-free motion capture can get the motions without the special suits, and thus are highly applicable to motion-based contents, e.g. dance and sports. The Kinect v2 released by Microsoft has been applied to many games as it is inexpensive and capable of extracting motions in real time. Yet, it has many limitations in extracting motions with a single sensor, resulting from the joint occlusion and other challenges.

To address the challenges resulting from the occlusion of body parts, more sensors are used to minimize the occluded parts. Lately, methods of using multiple Kinects have been suggested. Zhang and colleagues tracked postures with particle filtering and partition sampling [1]. Their method drew upon not skeleton data but template matching to estimate postures through optimization. Kitsikidis et al. used three Kinects to retrieve dance motions, and notably used HCRF to recognize motion patterns [2]. Kaenchan et al. analyzed walking motions based on the mean positions of joints tracked [3]. Moon et al. used the Kalman filtering to alter and mix accurate Kinect data [4]. Yet, they failed to capture 360-degree motion events because Kinects were placed in front and the motions were too simple. Jo et al. proposed a system using multi Kinects to track multiple users [5], but they focused on tracking the positions of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_5

multiple users instead of retrieving their motions. Ahmed used 4 Kinects to capture boxing and walking motions in 360° [6], tracked users' faces to determine a central Kinect with the joint inputs from the other Kinects being used to retrieve the joints that the central Kinect failed to track. Similarly, Baek et al. selected a central Kinect based on the movements of root joints and retrieved the postures by mixing the joints based on the weights of 5 segments tracked [7].

The present paper used 8 Kinects v2 to build a multiple Kinect system, and proposed a method of retrieving body motions from a series of noise joint data inputs from each sensor. The user motions were dynamic, e.g. Taekwondo, and could be captured in 360° in real time (30 fps). The proposed method captures dynamic motions with ease and fast without requiring any motion capture data or pre-trained probability model. The proposed multiple Kinect system was compared with a commercial motion capture system, Xsens to measure the accuracy of data recognized by the former.

2 Multi Kinects System and Data Transmission

As a single Kinect v2 can be connected to a single PC, N Kinects need be connected to N PCs. As in Fig. 1b, 2 Kinects were installed on all sides (front, rear, left and right), adding up to 8 Kinect systems. To incorporate the data inputs from each Kinect, the server-client model as in Fig. 1a was used, where N PCs were connected to the server PC. Upon being connected with the clients, the server sends the background removal command to the clients, where the backgrounds and noises are removed from the depth data to transmit the data specific to the body (i.e. the depth and joint data whose color values are mapped).

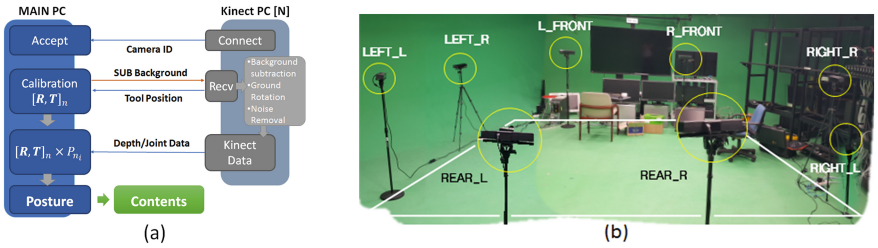


Fig. 1. System overview, (a) server-client model, (b) system configuration

The depth image that constitutes the background is saved when the initial Kinect depth data are acquired. Once the Kinect sensor senses the user, it compares the depth value of the P_{ij} pixel with that of B_{ij} pixel in the background saved. When the former is below the threshold, it is considered as the background, and excluded. As the depth data tend to show noises at the edges, the depth values of 8 pixels adjacent to the P_{ij} pixel are compared to determine the similarity of depth values. When the depth similarity is below k, it is considered a noise and thus excluded. Finally, as the floor around the user still includes the noise owing to the depth data, the depth data below a certain

value of user's ankle joint data are considered noises and thus excluded. Here, as the Kinect sensor could be tilted, the normal value of the floor is determined based on the vector associated with the root and spine-mid joints while the user stands upright in the initial setting. The gradient is corrected by calculating the rotation matrix, where the normal vector for the floor is matched to the up vector (0, 1, 0).

3 Calibration

As the coordinate systems of the data inputs from each Kinect differ from one another, they need be unified into a single coordinate system. Here, the front Kinect is selected as the reference coordinate system. For calibration, a long thin stick (about 50 cm) with a light cubic object (for recognition) at its tip is used as a tool. Based on the resolution of the Kinect, the depth data of the stick are ignored, while the depth data of the cubic object at the tip are taken. As the user moves the tool in the capture space, the mean value of the depth data of the object at the tip is saved as the central point. The user can set the timing and number of data to be captured. Here, 300 data are collected at an interval of 50 ms (Fig. 2). Excluding the data occluded by the body, the rotation matrix (R) and the translation vector (t) are calculated by applying the ICP (iterative closest point) technique to the points from the Kinects corresponding to the input points from the reference Kinect. As data are collected at a certain interval of time, the input point matching the central point is easy to find, which is conducive to fast and accurate calculation. Figure 2 shows the data inputs from the multiple Kinects prior to the calibration and the rotation and translation of the points after the calibration.

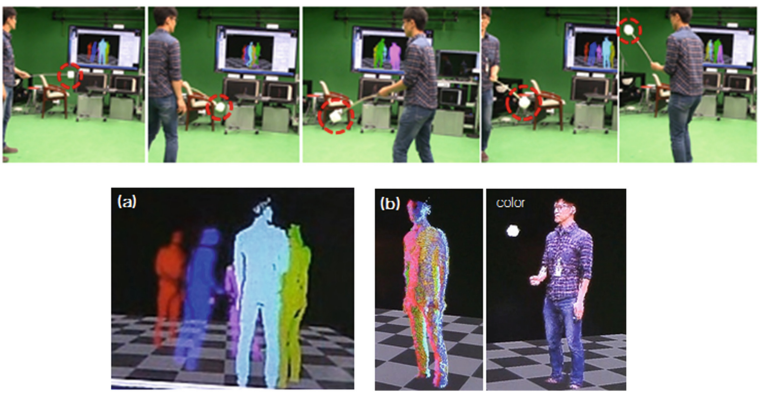


Fig. 2. Calibration process using the tool (Up), Calibration (a) before and (b) after (Down)

4 Joint Selection

The upgraded Kinect v2 enhances the accuracy of joint data over the existing version when the user faces forward. As in Fig. 3a, when the arms are lifted forward, the existing version tracks the elbow and wrist joints even in '*Not Tracked*' setting, calculates wrong positions, and is prone to errors. As in Fig. 3b, when the user turns right, it is impossible to track the positions because the right hip is blocked. Thus, the SDK 2.0 version continues to track wrong joint positions. Therefore, significant noises could occur due to the wrong joint positions when joint values are simply added up to determine mean positions or weighted. This challenge should be addressed to retrieve postures. In particular, the selection of root and hip joints when the user turns is most challenging.

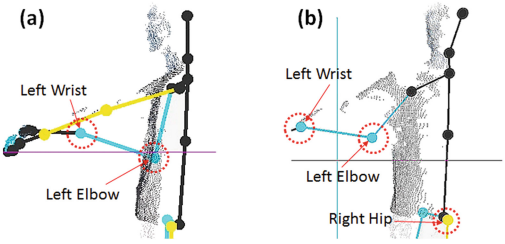


Fig. 3. Example of tracking error: (a) Arms are lifted forward, (b) User turns right

The present paper builds a model based on the user's initial posture and proposes a method of choosing the optimal joints for each part. The initial posture model is generated with the user standing with both arms spread wide while facing the front Kinect. At the same time, the length of each joint is measured. The initial model becomes the reference model, which is used to determine the left and right sides of the joints in the postures following. The lengths of joints may vary with the noises arising in the process of retrieving the posture. Thus, the reference model is used to correct the variation of joint lengths. As the Kinect does not tell left from right, the left joint seen from the front may be tracked as the right joint by another Kinect. Therefore, the calculation varies with whether it is necessary to distinguish left from right in retrieving a posture.

To retrieve the joints, the top nodes (root and hip joints) are first located. As aforementioned, it is not easy to find the accurate position of the hip joint because of lots of noises arising when the user rotates. In generating the initial model, not only the distance between the root and hips but also that between the hips should be measured. The closest values to the triangle (*LHip-Root-RHip*) measured in the initial model are found for the root and hip joint inputs from each Kinect. Based on the ratios of joint lengths, the joints whose values are below the given values are chosen. It is most likely that the data from the Kinect sensors facing the user and those placed in the rear are selected. Usually, up to two candidates are chosen, weighted based on similarities and mixed.

As the torso joints have no left and right sides clearly separated, mean values are used to calculate the joint positions, which are in turn adjusted based on the normal vectors associated with parent joints and the initial joint lengths.

As for the arms and legs, the K-means algorithm is used and the joint data are divided into two parts (S_A and S_B), each of which is to become the left or right side. The arms start from the top nodes, or the shoulder joints. The minimal-difference pattern in the square values of the distance between the previous posture's shoulder joint positions(Lj_{fr-1}, Rj_{fr-1}) and the two-part data's mean joint positions (P_A, P_B) is used to determine the left and right arms. Likewise, the elbows/knees and the wrists/ankles are calculated based on the differences in distances. Still, the parent joints serve as the references for comparing the joints.

As aforementioned, given the mixed values found by weights are significantly affected by noises, the selection is made based on the variation of joint angles. As for the reference for joint selection, the joints (J_s), which correspond to the minimal sum of the vector rotation direction (D_s) and rotation angle (A_s) calculated from the joint positions ($Pj_{\text{fr-2}}$, $Pj_{\text{fr-1}}$) and current joint position (Pj_{fr}), are selected.

Kinect finds the body parts based on learning data but sometimes fail to yield the joint values especially when feet go higher than the lower back as in kick motions. When no joint data are gained, the joint vector generated in previous postures (the vector between parent and children joints) is used to retrieve the posture.

5 Results

The present paper proposes a multiple Kinect system that captures motions in real time by minimizing the joint occlusion. As in Fig. 4, the user's posture can be retrieved although the hand is blocked or when the user rotates in 360°. Also, the proposed system can capture dynamic motions, e.g. Taekwondo.

Here, the proposed system is compared with the commercial motion capture system, Xsens to determine its accuracy. Xsens' data are saved as 120 and 240 frames. To match the initial setting, a T-pose is taken first. Motions are converted by matching the joint scales between Xsens' data and the multiple Kinect's data. To synchronize with

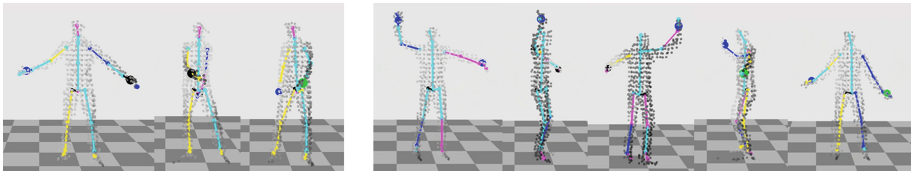


Fig. 4. Motion capture result: arm joint occluded by body (left), 360-degree left turn (right)

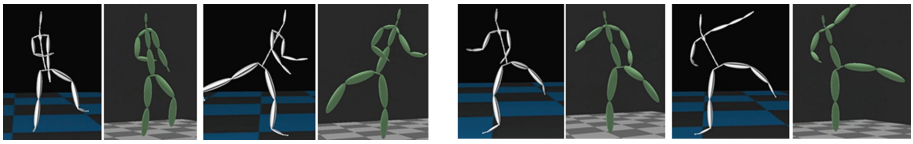


Fig. 5. Comparison of the motion: Xsens data (left-white), multi Kinects data (right-green)

the 30-fps Kinect data, the multiple Kinect system saves the data together with the timing of captures (milliseconds). The Xsens' data saved at a time closest to the recorded time are compared. The angular variation of joints from *N* data is calculated to compare the accuracy based on the difference in variation.

Figure 5 compares the data of two motions synchronized. 6 dynamic motions are measured in terms of the similarities of postures. The similarities are found as in Table 1. The lower extremity is less accurate than the upper one, because noises arise in the sensors attached to the feet in Xsens, and because errors occur in the lower extremity as the sensors are placed a bit high to increase the recognition of kicks in the multiple Kinect system. In particular, some motions such as the jump kick are not recognized by the multiple Kinect system, resulting in significant noises and errors (Fig. 6). Future research will draw upon the depth data to develop the technology for correcting the joint positions and for removing noises associated with legs and thus to increase the overall accuracy.

Table 1. Similarities of postures: six dynamic motions (M1–M6)

	Right-Arm	Left-Arm	Right-Leg	Left-Leg	Average
M1. 2,572 frames (240 Hz)	90.2 %	91.7 %	81.2 %	83.1 %	86.6 %
M2. 2,072 frames (240 Hz)	90.3 %	92.5 %	79.2 %	82.0 %	86.0 %
M3. 4,371 frames (120 Hz)	91.0 %	91.6 %	78.2 %	78.1 %	84.7 %
M4. 7,518 frames (120 Hz)	91.1 %	91.9 %	79.2 %	79.2 %	85.4 %
M5. 10,060 frames (120 Hz)	89.0 %	90.0 %	79.0 %	78.9 %	84.2 %
M6. 12,986 frames (120 Hz)	90.0 %	90.9 %	79.6 %	80.0 %	85.1 %
Total 39,579 frames	90.3 %	91.6 %	79.4 %	80.2 %	85.3 %

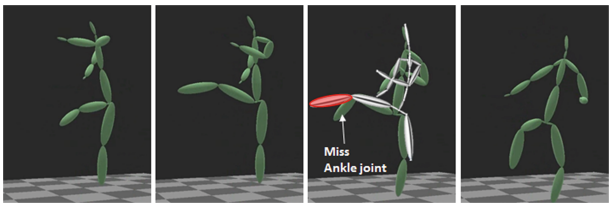


Fig. 6. Tracking failure of the ankle joint in the jump kick

Acknowledgments. This research project was supported by the Sports Promotion Fund of Seoul Olympic Sports Promotion Foundation from Ministry of Culture, Sports and Tourism.

References

1. Zhang, L., Sturm, J., Cremers, D., Lee, D.: Real-time human motion tracking using multiple depth cameras. In: IEEE/RSJ IROS 2012, pp. 2389–2395 (2012)
2. Kitsikidis, A., Dimitropoulos, K., Douka, S., Grammalidis, N.: Dance analysis using multiple kinect sensors. In: VISAPP 2014, Lisbon, Portugal, pp. 5–8, January 2014

3. Kaenchan, S., Mongkolnam, P., Watanapa, B., Sathienpong, S.: Automatic multiple kinect cameras setting for simple walking posture analysis. In: ICSEC 2013, pp. 245–249 (2013)
4. Moon, S., Park, Y., Ko, D.W., Suh, I.H.: Multiple kinect sensor fusion for human skeleton tracking using Kalman filtering. Int. J. Adv. Robot Syst. **2016**, 1–10 (2016)
5. Ahmed, N.: Unified skeletal animation reconstruction with multiple kinects. In: Eurographics 2014, pp. 5–8 (2014)
6. Baek, S., Kim, M.: Dance experience system using multiple kinects. Int. J. Future Comput. Commun. **4**(1), 45–49 (2015)
7. Jo, H., Yu, H., Kim, K., Jung, H.S.: Motion tracking system for multi-user with multiple kinects. IJUNESST **8**(7), 99–108 (2015)

Cell-Based Indexing Method for Spatial Data Management in Hybrid Cloud Systems

Yan Li and Byeong-Seok Shin(✉)

Department of Computer Information and Engineering,
Inha University, Incheon, Korea
leeyeon622@gmail.com, bsshin@inha.ac.kr

Abstract. In order to efficiently support various spatial and non-spatial queries over geographic heterogeneous cloud environments, we propose a cell-based inverted list index method. Our proposal includes a spatial keyword cell structure for simultaneously managing spatial and non-spatial keywords. An extended inverted list is constructed in order to support robust indexing of loosely coupled collections of heterogeneity spatial objects; therefore, our method can support flexible queries efficiently, such as keyword spatial and non-spatial queries and nearest neighbor queries. Experiment results show that the proposed indexing method can support quick answer of spatial queries compared with several typical existing indexing methods.

Keywords: Cell-based index · Inverted list · Keyword query

1 Introduction

Ubiquitous computing imposes new challenges that were already foreseen. Anywhere, anytime, and anything computing needs to cope with computing devices, users, and applications that are mobile [1].

Spatial, spatio-temporal, and other non-spatial unstructured data, such as images, exist. Based on heterogeneous data, flexible queries, such as keyword, time-interval, and nearest neighbor, are required. An example of such data is, "Return the cars that passed by the white building whose names contain 'high-tech' from 9:00 am and 9:02 am." How to index such heterogeneous data and process their flexible queries is a significant challenge.

Many studies have been conducted on developing strategies for indexing and querying spatial and spatio-temporal data. R-tree based index methods present spatial objects using MBR (Minimum Bounding Rectangle) [2]. HR-tree introduces timestamps into R-tree aimed at managing spatial-temporal data. It constructs current independent R-trees for each timestamp. 3DR-tree [3, 4] treats time as yet another dimension, in addition to the spatial dimension, and it can also support both spatial and spatial-temporal objects. MV3R-tree [5, 6] is based mainly on the multi-version B-tree and builds two trees to support timestamps and interval queries. Almost all the existing R-tree-based methods can only manage spatial and spatial-temporal data separately [7, 8], and cannot manage the queries required for both types of data. In addition, existing methods only focus on spatial and temporal information, and cannot manage non-spatial attributes [9–11].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_6

In this paper, we propose a novel indexing method, called cell-based inverted list indexing method for graphical heterogeneous data. It introduces an inverted list to manage both spatial and spatial-temporal data and support various queries. The proposed index, called SOL (Spatial Object List), is an extended inverted list used to support robust indexing of loosely coupled collections of heterogeneous spatial objects, and when it is extend to the time dimension, it can manage spatial-temporal objects. This indexing method manages geographic heterogeneous data, both spatial and temporal information, and it can support flexible queries, such as keyword, time slice, time interval, and nearest neighborhood.

2 Cell-Based Inverted List Index Method for Hybrid Spatial Data

For the spatial data in GIS (Geography Information System), geographic entities are physically continuous entities with static positions in the geographic space where people can take action. Shops are geographic entities, as are stations, buildings, sightseeing spots, and landmarks. In general, all spatial data models fall into two basic categories, as described in the following paragraph.

The first category is a vector data model with discrete features, such as building locations, pole locations, postal code areas, and lakes that can be summarized by areas usually represented using the vector model. The second category is a raster data model with continuous numeric values. Several continuous categories, such as elevation, soil type, and rainfall, are usually represented using the raster model.

In this paper, we only consider spatial data with discrete features, such as buildings and mountains. This type of spatial data mainly appears in spatial and non-spatial data. While implementing GIS, the most common sources of attribute data are an organization's databases combined with data sets bought or acquired from other sources to complete gaps.

The triples in our method correspond to instances extracted from the data source using a variety of methods. For example, we extract entries from tuples in a relational database by attempting to guess the E/R model that can lead to the schema. Other examples include adapting techniques from previous works to extract relationships from both structured and unstructured data. Note that these extractions are imprecise in nature, and thus our querying mechanisms and indexing techniques need to allow more flexibility.

The current indexing methods build a separate index for each data source in order to support structured queries on structured data, and they can only manage location information. For the non-spatial and unstructured data contained in geographic data, inverted lists are the main method created in other existing methods to support keyword queries. Consequently, as we demonstrate, they are limited not only in their treatment of both location and non-spatial information, but also in the context of queries that combine structure and keywords.

The keywords in inverted lists are ordered alphabetically, and the instances are ordered by their identifiers. Table 1 demonstrates how to use inverted lists that represent instances. Here we add the attributes as part of the keywords, and the form of the

Table 1. Inverted list for a set of instances

	A	B	C	D	E
Blue//Color	0	0	0	0	1
Backgate//Name	0	0	1	0	0
Coffee//Name	0	0	0	0	1
Hana//Name	0	1	0	0	0
Hightech//Name	1	0	0	0	0
Main//Name	0	0	0	1	0

keyword is "keyword//attribute". If the cell is not zero, we state that the corresponding instance at that column is indexed on the corresponding keyword at that row.

Note that inverted lists, as described in Table 1, do not capture any location information. Furthermore, SOL is the indexing method that combines grid cells with inverted lists. Conceptually, SOL is a two-dimensional table, where the i_{th} row represents the keyword element that contains the keyword and attribute, and the j_{th} column represents the location information. Cell (i, j) is a pointer to a list that stores a set of spatial objects.

Here, SOL adopts an overlapping technique. For some instances that receive several cells, the pointers of those cells point to the same instance. In order to save disk space, only one instance is stored and the pointers of that instance point to the same list that stores the instance. Hence, pointers 131 P and 132 P in Fig. 1 point to the same list that contains instance E.

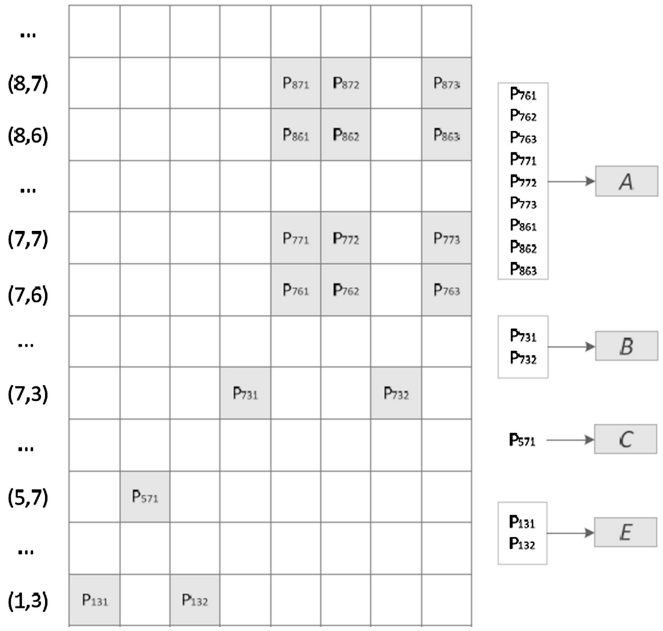


Fig. 1. Spatial object list for several instances

The Spatial-Temporal Object List (STOL) is a collection of Spatial-Temporal Object Tables (STOT), each of which serves as an index for a set of moving objects at a specific time interval. At specific time t, spatial-temporal data can be modeled with grid cells in the two-dimensional $X \times Y$ space in the form (x, y). Unlike this, in our method, we map the two-dimensional (x, y) into a one-dimensional L taken as the y-axis, and we take the pointers of the spatial-temporal objects as the x-axis. If the value of each cell is not zero, this means that the spatial-temporal object appears at the corresponding location at that time. One of the fundamental objectives of the proposed method is to efficiently manage various queries, such as keyword, keyword nearest neighborhood, time-interval, and time-slice. The structure that extends the cell-based structure using an inverted list makes the query processing for keyword and keyword neighborhood queries intuitive and straightforward. Time-interval and time-slice query processing involves selecting a set of entries that correspond to the timestamp interval in the query. The cell-based structure of STOL is good for time-interval and time-slice queries. Here, we provide several examples to explain how to process those queries.

3 Experimental Result

In our performance evaluation, we mainly use TIGER/Line shape files [12] and the data sets generated by the GSTD spatio-temporal generator [13]. GSTD has been widely employed as a benchmarking environment for access methods that manage moving points. Each of the following data sets contains 10,000 regions with density 0.5 and is generated as follows: the initial positions of the objects are determined following a Gaussian distribution. Timestamps are modeled as floating numbers that range from 0 to 1 with granularity 0.01. The performance of different access methods is measured by running workloads. Each workload contains 500 queries with the same area and interval length.

Because there are no current methods that manage both spatial and moving objects (to the best of our knowledge), we divided our experiments into two parts: one is responsible for comparisons with Hybrid-ATIL, a heterogeneous data indexing method; the other part compares our method with STR-tree, HR-tree, 3DR-tree, and MV3R-tree—all of which are R-based indexing methods.

Figure 2 shows the average results of building time, memory usage, keyword query, and keyword neighborhood query execution time. It shows that the Hybrid-ATIL method requires slightly longer time to process the two types of queries because it stores both the attributes and their association. The index lookup time of our method requires less time than that required by Hybrid-ATIL on both keyword and keyword neighborhood queries.

In the second set of experiments, we compared our method with several R-tree-based indexing methods. We divided the data space into 64 cells and used the data sets with cardinality 10 K that evolved for 624 timestamps. The number of different types of queries is 100, and they are range, time-interval, and time-slice queries. Using a page size of 1 KB, the fanouts of those R-tree based indexing methods are between 28 and 36.

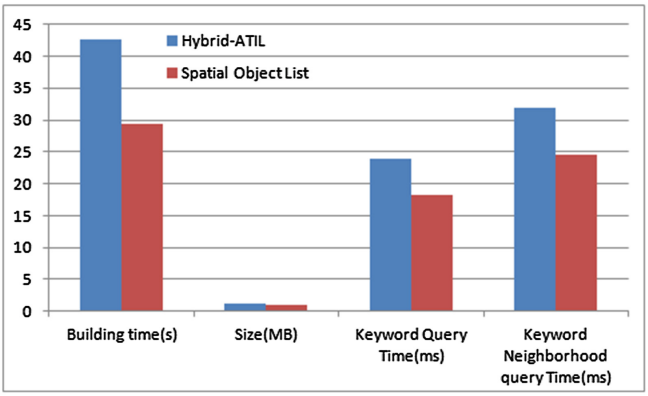


Fig. 2. Results of comparison with Hybrid ATIL (construction time, memory usage, and query run time)

Figure 3 shows the average time for processing different types of queries. For all indexing methods, the range query required the longest time. Furthermore, for the time interval and time slice queries, our method outperforms the R-tree-based methods.

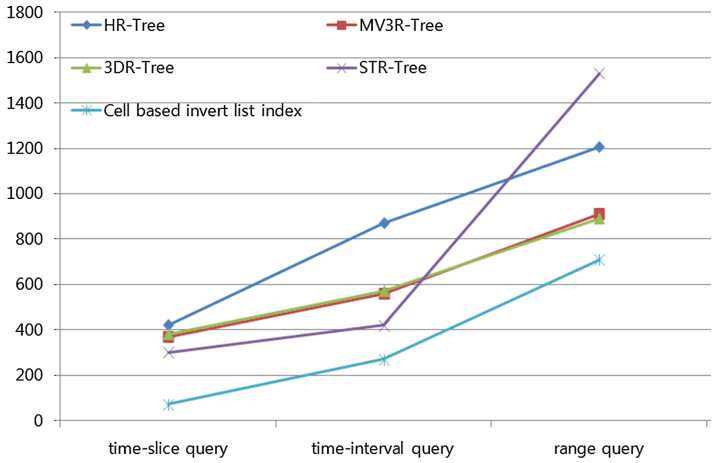


Fig. 3. Average response time for three types of queries

4 Conclusion

In this paper, we introduced a cell-based inverted list index method that uses inverted lists to extend cell-based indexing structures in order to overcome the limitations of hierarchical access methods. SOL was used for indexing spatial data and STOL was used for indexing spatial-temporal data. We also provided the definitions of several types of queries over heterogeneous geographical data, which are keywords and

keyword neighborhood queries, as well as those queries that combine keyword and structured queries. Our indexing method can support hybrid spatial data management of keyword queries, and we also provided a set of algorithms for query processing. Through a set of experiments, the proposed method proved to be an effective indexing method well suited for managing spatial and non-spatial data.

Acknowledgments. This research was supported by a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant number HI14C0765).

References

1. Guttman, A.: R-trees: a dynamic index structure for spatial searching. In: Proceedings of the International Conference on Management of Data, pp. 47–54. ACM Press (1984)
2. Yannis, T., Michael, V., Timos, S.: Spatio-temporal indexing for large multimedia applications. In: Proceedings of the International Conference on Multimedia Computing and Systems. IEEE Press (1996)
3. Tao, Y., Papadias, D.: MV3R-tree: a spatiotemporal access method for timestamp and interval queries. In: Proceedings of the 27th International Conference on Very Large Databases, pp. 431–440. Morgan Kaufmann Publishers Inc., San Francisco (2001)
4. Mokbel, M.F., Ghanem, T.M., Aref, W.G.: Spatio-temporal access methods. IEEE Data Eng. Bull. **26**(2), 40–49 (2003)
5. Markov, K.: Multi-dimensional context-free access method, Ph.D. thesis, Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia (2006)
6. Markov, K., Ivanova, K., Mitov, I., Karastanev, S.: Advance of the access methods. Inf. Technol. Knowl. **2**, 123–137 (2008)
7. Mario, A.N., Silva, J.R.O.: Towards historical R-trees. In: Proceedings of the 1998 ACM Symposium on Applied Computing, Atlanta, GA, pp. 235–240, February 1998
8. Tao, Y., Papadias, D.: MV3R-tree: a spatio-temporal access method for timestamp and interval queries. In: Proceedings of 27th International Conference on Very Large Data Bases, Roma, Italy, September 2001
9. Choi, W., Moon, B., Lee, S.: Adaptive cell-based index for moving objects. Data Knowl. Eng. **48**, 75–101 (2004)
10. Jiang, H., Lu, H., Wang, W., Ooi, B.C.: XR-tree: indexing XML data for efficient structural joins. In: Proceedings of ICDE (2003)
11. Aung, S.N., Sein, M.M.: Hybrid geo-textual index structure for spatial range keyword search. Comput. Sci. Eng. **4**(5/6), 21 (2014)
12. Thompson, J.H., Blumerman, L.: 2015 TIGER/Line Shapefiles (machine-readable data files) Technical Documentation, U.S. Census Bureau (2015). http://www.census.gov/geo/www/TIGER/TIGERua/ua2ktgr.pdf
13. Theodoridis, Y., Silva, J.R.O., Nascimento, M.A.: On the generation of spatiotemporal datasets. In: Güting, R.H., Papadias, D., Lochovsky, F. (eds.) SSD 1999. LNCS, vol. 1651, pp. 147–164. Springer, Heidelberg (1999). doi:10.1007/3-540-48482-5_11

SOA Based Equipment Data Management System for Smart Factory

YunHee Kang[1], Soong-ho Ko[2], and Kyoungwoo Kang[1(✉)]

[1] Division of Information Communication, Baekseok University, 115 Anseo-Dong, Cheonan, Chungnam 330-704, Korea
{yhkang, kwkang}@bu.ac.kr

[2] DeviceEng Co., Ltd., 169 Eumbong-Ro, Eumbong-Myeon, Asan-si, Chungnam 336-864, Korea
ksh@deviceeng.ac.kr

Abstract. In this paper we introduce a SOA based EDA system complied with SEMI Standards for smart factory in semiconductor manufacturing filed To design the system that is used to integrate information systems in the factory, we analyze the requirements of EDA system in the prospect of the components of the information systems. We build a prototype for an EDA system including EDA Host and EDA Client in EDA SEMI standards.

Keywords: SOA · EDA · Smart factory · SEMI standards

1 Introduction

To collect and use data from the equipment in the factories, the SEMI Standards have evolved to support the demand for data exposed by the process analysis and control applications. The process used to manufacture semiconductors is highly complex and requires advanced manufacturing equipment, highly controlled environments, and the use of specialized chemicals to produce these complex products [1].

Hence many applications of manufacturing factories require variety and complex data types from integrated information framework [2, 3]. EDA (Equipment Data Acquisition) can be used to get significantly higher trace data collection throughput, and the robust tool model in EDA provided better access to sensors and other key equipment variables useful for operational data monitoring [4].

In this paper we introduce a SOA based EDA system complied with SEMI Standards for smart factory in semiconductor manufacturing filed. To design the system that is used to integrate information systems in the factory, we analyze the requirements of EDA system in the prospect of the components of the information systems such as ERP, SCM, PLM, etc. We build a prototype for an EDA system including EDA Host and EDA client in EDA SEMI standards [7–10].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_7

2 Related Works

The EDA standard is announced for the semiconductor industry to improve and facilitate communication between data gathering software applications and the factory equipment each other effectively. It is a suite of model-based communication standards that enables devices of arbitrary size and complexity to securely publish data to multiple, distributed clients using standard internet technology. The main EDA SEMI standards include E120, E125, E132, and E134 [7–10]. Solutions must comply with the specific SOAP (Simple Object Access Protocol)/XML(Extensible Markup Language) implementations of these standards; E120.1, E125.1, E132.1, and E134.1. Figure 1 shows the overall EDA standards.

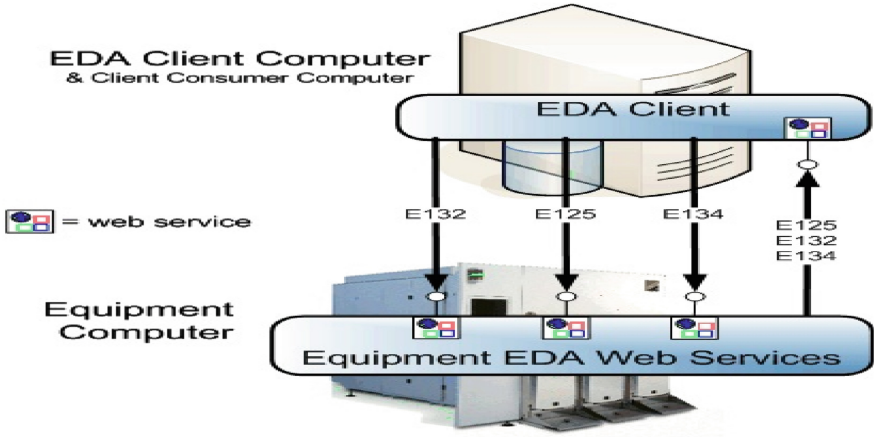


Fig. 1. Overview of EDA standards

3 Design of EDA Architecture

3.1 Experimental Setup

To study the performance of factory scale EDA in Ethernet interface, we design the EDA system that consists of EDA client and EDA Host. It is in conjunction with a testbed composed of PLC associated devices, EDA Host and EDA client. EDA client has several applications like EDA configurator and equipment data collector. Service-oriented architecture (SOA) presents a fundamental shift in dealing with the difficulties of building distributed systems [5, 6]. SOA presents a fundamental shift in dealing with the difficulties of building distributed systems. Figure 2 shows the overall architecture for building an EDA application of the equipment in a semiconductor process.

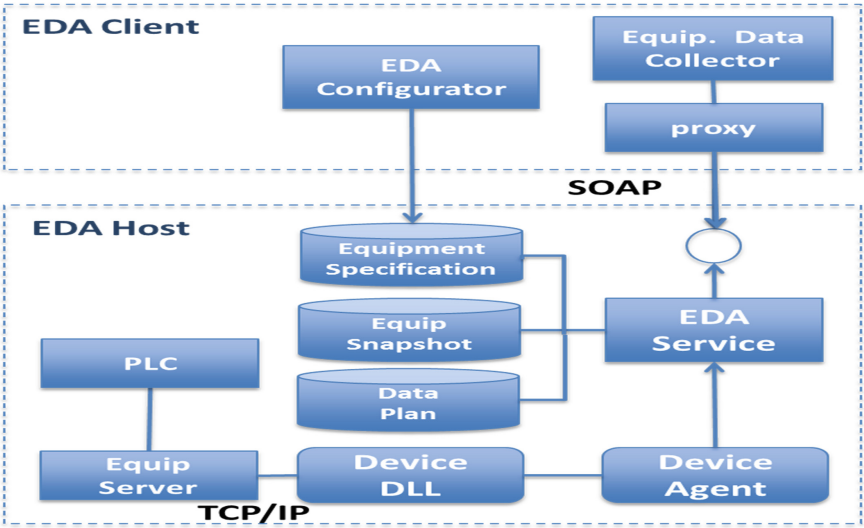


Fig. 2. The overall EDA structure

This testbed is a C# with WCF of the EDA communication infrastructure [4]. WCF is a framework for building service-oriented applications and a software development kit based on .NET for services on Windows and is a unified programming model for building service-oriented applications. It is used for clients and services to send messages between each other. Its services are interoperable, which uses a variety of network protocols such as HTTP, TCP, MSMQ, etc. Services are offered on an endpoint. WCF separates service hosting, endpoints and services. In WCF, all services expose contracts. The contract is a platform-neutral and standard way of describing what the service does. WCF defines four types of contracts. Applications or services are designed for distributed environment using a single model. The testbed recreates network traffic expected from a real world, factory scale EDA implementation, including various data types such as; exception reports, event reports and trace reports. Further, the testbed can be scaled up to simultaneously.

3.2 Experiments Result

To evaluate EDA testbed, we apply it to the AutoFOUP cleaning system that is particularly well suited to the precision cleaning of the 300 mm (12”) wafer container, FOUP. Figure 3 shows the components of AutoFOUP that consists of buffers, chambers, transfer robot and ports.

Basically equipment metadata is supplier-sensitive information. It can be access via an authenticated session with the equipment. In E125, metadata interface is well modular written in WSDL. The `EquipmentMetadataManager` provides 9 methods including `GetTypeDefinitions`, `GetEquipmentStructure`, `GetEquipmentNodeDescriptions`, etc. To get the overall structure of a equipment, an EDA

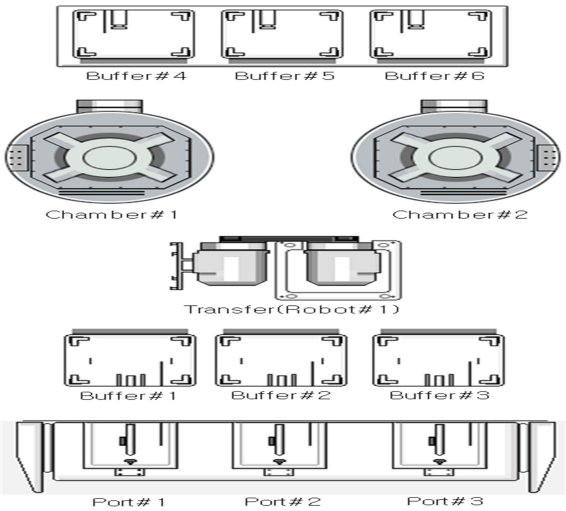


Fig. 3. The components of AutoFOUP

client call a method `GetEquipmentStructure` in `EquipmentMetadataManager`. The following is the element of `GetEquipmentStructure` method in a WSDL:

```
<wsdl:portType name="EquipmentMetadataManager">
  …
  <wsdl:operation name="GetEquipmentStructure">
    <wsdl:input name="GetEquipmentStructureIn"
        message="eqSDPort:GetEquipmentStructureRequestMessage"/>
     <wsdl:output name="GetEquipmentStructureOut"
       message="eqSDPort:GetEquipmentStructureResponseMessage"/>
     </wsdl:operation>
  …
</wsdl:portType>
```

The response of this call GetEquipmentStructureResponseMessage provides a structure of a specific equipment predefined as CEM(Common Equipment Model) in E120 within a SOAP message. Figure 4 shows the SOAP message for a method GetEquipmentStructure with input and output messages.

The element Equipment represented with an XML schema complied with E120 is as followed. It gives keys that are used for searching details about the equipment including subsystems, IO device, and modules by XPath. The element Equipment is associated with an element type `EquipmentType` represented with complex type in XML schema. In WCF programming model, applications have unique needs for exposing functionality through service contracts and for choosing the right mechanism for serializing complex types for each service operation. A service has contracts. EDA client gets the metadata embedded in a SOAP reply message complied with E120

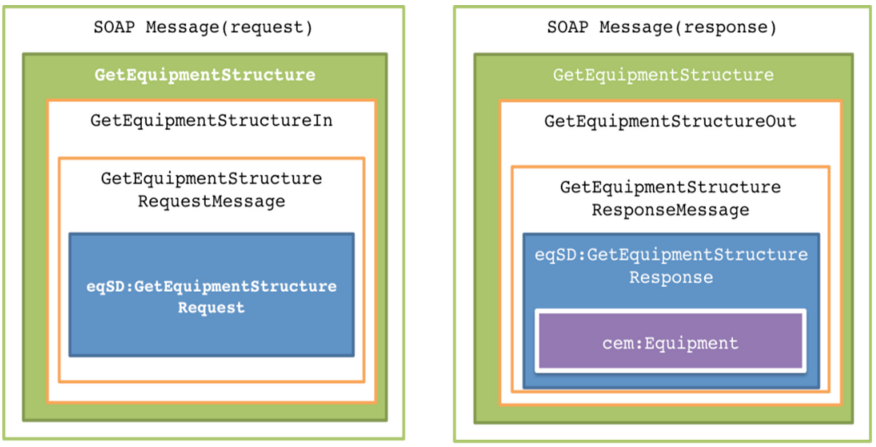


Fig. 4. The SOAP message structure for `GetEquipmentStructure` method

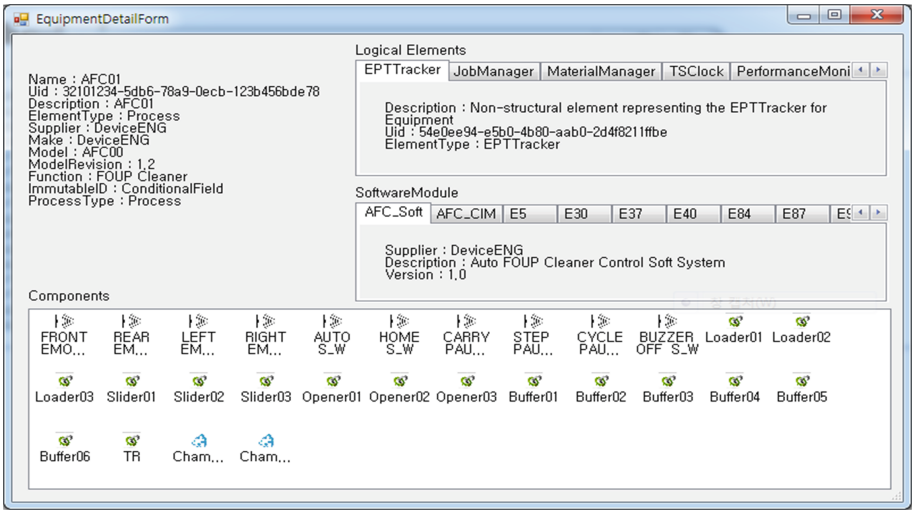


Fig. 5. The detail information about AutoFOUP

standard from EDA host. Figure 5 shows the description of AutoFOUP equipment as the result of `GetEquipmentStructure`.

4 Conclusion

In this paper we introduced an EDA system complied with SEMI Standards for smart factory. Hence we analyzed the requirements of EDA system in the prospect of the components of the information systems. To show the main operations of EDA based on

web services, we designed a SOA based EDA system for semiconductor manufacturing. It is useful to integrate heterogeneous service platforms and diverse equipments in a factory environment. As the result, standard techniques of web services like SOAP and WSDL give advantage to transform the traditional information systems with low level messaging to well-designed systems to optimize operation of manufacturing in smart factory.

Acknowledgements. This work (Grants No. C0395837) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2016.

References

1. May, G.S., Spanos, C.J.: Fundamentals of Semiconductor Manufacturing and Process Control. Wiley, New York (2006)
2. Zuehlke, D.: Smartfactory–from vision to reality in factory technologies. IFAC Proc. **41**(2), 14101–14108 (2008)
3. Lucke, D., Constantinescu, C., Westkämper, E.: Smart factory - a step towards the next generation of manufacturing. In: Mitsuishi, M., Ueda, K., Kimura, F. (eds.) Manufacturing Systems and Technologies for the New Frontier: The 41st CIRP Conference on Manufacturing Systems, pp. 115–118. Springer, London (2008)
4. Li-Baboud, Y.S., Zhu, X., Anand, D.M., Hussaini, S., Moyne, J.R.: Semiconductor manufacturing equipment data acquisition simulation for timing performance analysis. In: Proceedings of IEEE ISPCS (2008)
5. Erl, T.: Service-Oriented Architecture: Concepts, Technology, and Design. Prentice Hall PTR, Upper Saddle River (2005)
6. Spiess, P., Karnouskos, S., Guinard, D., Savio, D., Baecker, O., Souza, L.M.S.D., Trifa, V.: SOA-based integration of the internet of things in enterprise services. In: IEEE International Conference on Web Services, ICWS 2009, pp. 968–975, Los Angeles, CA, USA (2009)
7. SEMI E120-1.V1104 specification for the common equipment model
8. SEMI E125-1.V0305 specification for equipment self description
9. SEMI E132-1.V0305 specification for equipment client authentication and authorization
10. SEMI E134-1.V1105 specification for data collection management

The Problem Analysis of Specific Personal Information Protection Assessment in Japan Case

Sanggyu Shin(✉), Yoichi Seto, Mayumi Sasaki, and Kei Sakamoto

Advanced Institute of Industrial Technology, 1-10-40, Higashiooi, Shinagawa-Ku, Tokyo 140-0011, Japan
{shin, seto.yoichi}@aiit.ac.jp

Abstract. In this paper, we analyze the total item assessment reports that have been published by municipalities for the mandated implementation of the specific personal information assessment in three perspectives. The three perspectives are (1) Adequacy of risk items, (2) Re-use of the assessment report, and (3) Classification of the assessment model. As a result, for example, in risk measures where there are many assessment reports, there is a description of the measures in the system but there are missing measures outside the system such as operation, etc.

Keywords: Specific personal information protection assessment · Specific personal information · Assessment report · My number

1 Introduction

Residents in Japan were notified about the "My number" system on October 2015. The personal information including my number is called "Specific personal information."

Implementation of the "Specific Personal Information Protection Assessment," which we also refer to "Protection Assessment" has been required in the appropriate municipal offices to keep certain personal information [1].

Protection assessment is done to prevent infringement of privacy of personal information and ensure the trust and protect the rights of citizens and residents. After protection assessment, each municipality must conduct their risk assessment.

The results of the protection assessment are published as an "Assessment report." However, it has been pointed out that the protection assessment may not have been properly implemented [2].

In this paper, we analyzed the report published by the municipality in the following perspectives: (1) Adequacy of risk items; (2) Re-use of the Assessment report; and (3) Classification of the Assessment model [3].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_8

2 Overview and Issues of Specific Personal Information Protection Assessment

Protection assessment is classified into three aspects of evaluation by the threshold decision: basic items assessment, priority items assessment, and all items assessment. Threshold decision is affected by the number of target people, the number of transactors, and the occurrence or non-occurrence of major accidents of specific personal information.

The case which puts evaluation of all items into effect treats a lot of specific personal information more than other evaluation. Also, since there is a large number of persons handling the information, there is a high risk for leakage of specific personal information and other accidents. Therefore, all item assessment report (From now on referred to as the "Assessment Report") is necessary to evaluate concrete risk measures in more detail. In this paper, the assessment reports were analyzed in three aspects that have been pointed out by the persons concerned as targeted by the assessment report.

- **Adequacy of the Risk Item**
  The protection assessment, which is carried out by a municipality is performed to describe the contents of a risk measure to the risk item indicated on an evaluation document beforehand. However, the risk items are not uniform, and the standards used by municipalities when considering a risk measure are not specified in detail. Therefore, it is likely that there is a difference in the level of the methods and measures to select various things such as the municipality of risk items.
- **Re-use of the Assessment Report**
  It is not a big difference that the contents of office work are defined by law in the municipality, except partial for the municipality. Then, it is also conceivable to reuse the contents of the assessment report, which was evaluated previously in the same municipality.
- **Classification of the Assessment Model**
  In the office work that handles specific personal information, it is possible to perform an information link via the information provided by the network system. Therefore, the scope of protection assessment of municipality is asked to be evaluated, including the cooperation foundation such as the intermediate server of the relevant office works and the providing information network system, etc.

3 Analysis of Issues

3.1 Adequacy of Risk Items

In the protection assessment, assessment depends on the municipality because there is no procedure manual for risk evaluation. There is a possibility that proper implementation of risk evaluation is difficult because the person who estimates risk does is not specialized as we have also investigated in actual conditions. We target the all item assessment report for analysis because it puts risk analysis into all risk issues comprehensively. The 221 assessment reports exhibited were analyzed on (June 10, 2015) from

a specific personal information protection committee. The analysis is the same as office works, which targets the assessment report for the "Office works concerning the Basic Resident Register." Nine cases were analyzed, which corresponds to about 10 % of the assessment report of the target affairs that has been published at that time (80 cases).

We make a comparison between the assessment standard that we created and the assessment report of the municipality to be analysis target [4]. The result of the comparison is indexed in Table 1 to confirm the excess or deficiency for each corresponding risk item.

Table 1. The category of assessment of the risk response.

Assessment results	Assessment index
The risk correspondence indicated by the evaluation standard is being satisfied. Furthermore, the risk described corresponding to the evaluation criteria is supported.	3
The only parts of the risk management that are shown in the assessment criteria are described.	2
The risk correspondence indicated by the assessment standard isn't mentioned.	1
Risk correspondence isn't indicated in the assessment standard.	- (Excluded from assessment)

Table 2 shows the average value of the distribution and all the items of the assessment index for risk correspondence of system in the municipality. The assessment index when not mentioning the risk correspondence indicated by the assessment standard at all, is 1 point.

Average values of assessment index of all 49 items were conducted in municipalities is likely not to be applied appropriate assessment when close to 1 point.

Table 3 shows the distribution of the assessment index of the response to the management risk in municipalities and average values of the all item assessment.

Table 2. The situation of corresponding to the risk (System).

All 49 items	System				Assessment index (Average of all item)
	3	2	1	0	
A city	7	12	5	25	2.08
B city	11	10	5	23	2.23
C city	7	12	6	24	2.04
D ward	11	8	8	22	2.11
E city	9	13	3	24	2.24
F city	10	13	1	25	2.38
G city	11	12	1	25	2.42
H city	5	16	3	25	2.08
I ward	24	0	0	25	3.00

Table 3. The situation of corresponding to the risk (Management).

All 49 items	Management				Assessment index (Average of all item)
	3	2	1	0	
A city	11	17	7	14	2.11
B city	12	21	2	14	2.29
C city	9	16	11	13	1.94
D ward	12	12	12	13	2.00
E city	8	17	11	13	1.92
F city	12	14	10	13	2.06
G city	8	16	12	13	1.89
H city	11	13	12	13	1.97
I ward	36	0	0	13	3.00

3.2 Re-Use of the Assessment Report

The analysis target is personal confirmation information file on an assessment report. The number of characters to which an assessment report and a mentioning point are parallel is counted, and its ratio is calculated [5]. The indexing would make based on Table 4.

Table 4. Concordance assessment index.

Concordance rate	Concordance assessment index
Disagreement	0
Less than 25 %	1
More than 25 % ~ Less than 50 %	2
More than 50 % ~ Less than 75 %	3
More than 77 % ~	4

3.3 Classification of the Assessment Model

In the municipality, the My Number system promotes task collaboration with other government agencies by information cooperation. Information cooperation is performed by the intermediate server and providing information network system to be established. The intermediate server performs information cooperation and dissemination of information in the network system. The scope of the specific personal information protection assessment is expressed in four models, as illustrated in Fig. 1. The bold line frame of Fig. 1 illustrates the extent of the assessment, and the dotted box refers to an original specific personal information file and a duplicate DB. Table 5 shows the model classification of the assessment report of the municipality.

- Model A: The model that assesses relevant office work.
- Model B: The model that assesses relevant office work from the intermediate server.
- Model C: The model to assessment, including the intermediate server and information provided network system on relevance office work.
- Model D: The model to assessment separately the intermediate server which is assessment the information furnished the network system that relevance office work.

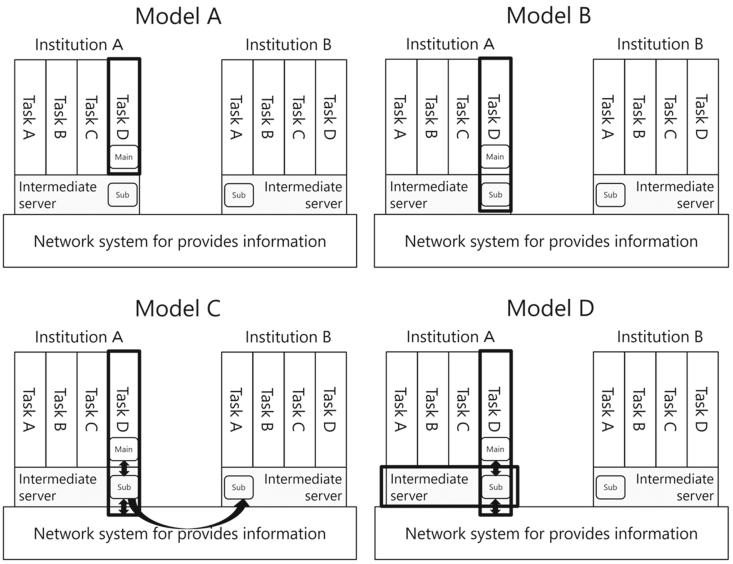


Fig. 1. Assessment Model

Table 5. The model classification of the assessment report.

Model	The subject and scope	Model classification of assessment report	Comments
A	Assignment task only	–	Not the object of evaluation such as an intermediate server. Assignment of the contents of an individual problem.
B	Assignment task ∼ Intermediate server	C city	Among the municipalities, it has published an evaluation report on the early (2014 November).
C	Assignment task ∼ Intermediate server ∼ Network system for provides information	A city, B city, D ward, E city, F city, G city, H city	Initially, the municipality of Model A there were many, been pointed out from a specific personal information protection committee, it was correct to Model C.
D	Assignment task ∼ Network system for provides information/Intermediate server	I ward	Has submitted in separate reports the evaluation and assessment of the intermediate server and office system.

Office systems are often packaged products and it is difficult for municipality officials to get familiar with the technical specifications particularly the detailed specifications of the intermediate server and information provided the network system. In addition, the risk in the system and the operation are the causes because of the measures are also fundamentally different, whose risks should be assessed in isolation.

4 Conclusion

We analyzed the implementation of the risk assessment based on published assessment reports as the target because it was pointed out that assessments may not have been adequately implemented. As a result, we came up with a description of risk measures in the system. However, many assessment reports such as system management are missing. Furthermore, the municipality assumed that office work was estimated through a target of evaluation, but when it lacked in knowledge to a system actually, it was revealed that there is a possibility that the system and the files, which were assessment targets were overlooked.

Acknowledgments. This research carried out in the Project Based Learning in the Advanced Institute of Industrial Technology. In advancing the PBL, we got the cooperation of Kazuhiro Midorikawa, Yuta Kurosawa, Okimura Seiji, and Xiaofei Ma. We would like to express our appreciation here.

References

1. The Specific personal information protection committee.: Description of the specific personal information protection evaluation guidelines. http://www.ppc.go.jp/files/pdf/explanation.pdf
2. Takashi, M.: Improper specific personal information protection evaluation shakes “My number” system. NikKei Computer, May 14 2014, pp. 6–10 (2015)
3. The Specific personal information protection committee.: My number protection assessment Web. http://www.ppc.go.jp/mynumber/evaluationSearch/
4. Yoichi, S.: The guideline of specific personal information protection assessment practice for the municipality. Gyosei, Kyoto (2015)
5. The specific personal information protection assessment book mentions point about the office work about the Basic Resident Register. http://www.ppc.go.jp/files/pdf/260624siryo1.pdf

Using a Fine-Grained Hybrid Feature for Malware Similarity Analysis

Jing Liu(✉), Yongjun Wang, Peidai Xie, and Xingkong Ma

College of Computer, National University of Defense Technology, Changsha, Hunan, China
liujing_nudt@nudt.edu.cn

Abstract. Nowadays, the dramatically increased malware causes severe challenges to computer security. Most emerging instances are variants of previously encountered malware through polymorphism and metamorphism techniques. The traditional signature-based detecting methods are ineffective to recognize the enormous variants. Malware similarity analysis has become the mainstream technique of identifying variants. However, most existing methods are either hard to handle polymorphic and metamorphic samples based on static structure feature, or time consuming and resource intensive by using dynamic behavior feature. In this paper, we propose a novel malware similarity analysis method based on a fine-grained hybrid feature by exploiting the complementary nature of static and dynamic analysis. We integrate dynamic runtime behavior with static function-call graph. The hybrid feature overcomes the limitation of using static and dynamic feature separately and with more accuracy. Furtherly, we use graph edit distance, and inexact graph matching algorithm as metric to measure the distance between malicious instances. We have evaluated our algorithm on real-world dataset and compared with other approach. The experiments demonstrate that our method achieves higher accuracy.

Keywords: Similarity analysis · Function-call graph · Hybrid feature · Graph edit distance

1 Introduction

Malware poses a major threat to network security. According to the latest report of Symantec, more than 430,000,000 new malware samples were discovered in 2015, up 36 percent from the year before. The sheer volume of malware brings severe challenges to security vendors. However, research shows that the majority of new incoming malware instances are merely variations of encountered malware through polymorphism and metamorphism techniques. They share the same functionality while have different syntactic representations.

Malware similarity analysis has been put forward to efficiently cope with the tremendous number of variants. Through precisely measuring the similarity based on quantitative metric to determine whether a malware program is similar to a previously-seen sample. A large amount of time and resources could be saved to avoid

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_9

the duplicated analysis of variants. It is the basis for automatic malware detection. It is also the foundation of malware classification and phylogeny model generation.

In this paper, we propose a novel malware similarity analysis metric using a fine-grained hybrid feature, which combines static function-call graph and dynamic runtime traces in a way that taking advantage of both simultaneously. Firstly, we extract the function-call graph of programs using static analysis, which is resilient to low-level obfuscation, such as basic block-reordering, register reassignment. Each vertex in function-call graph represents a function and each edge represents a caller-callee relationship between functions. Then we extract the dynamic runtime trace sets as function labels. At last, we use graph edit distance, an inexact graph matching method as metric to calculate the similarity degree among malicious samples. We have evaluated our algorithm on real-world dataset and compared with other approach. The experiments demonstrate that our method achieves higher accuracy.

The rest of this paper is organized as follows. We review the related work in Sect. 2. In Sect. 3, we describe the overview of framework. Then we introduce the extraction of the hybrid feature in Sect. 4 and the calculation of similarity metric is presented in Sect. 5. Section 6 evaluates the result of experiment. Finally, a summary of the paper is given in Sect. 7.

2 Related Work

Malware similarity analysis has attracted considerable attention. Most existing methods based on either static features or dynamic features.

Static features are extracted from malicious programs without executing it. Shafiq et al. [1] proposed to extract distinguishing features from portable executable (PE) format using the standard structural information that Microsoft Windows operating system defined. Kolter et al. [2] used n-grams of byte codes presented in the malware binary as features. Xin et al. [3] employed function-call graph which is a high-level structural feature for malware classification. However, as mentioned by Moser et al. [4], static analysis is difficult to handle the advanced obfuscation or self-mutating instances, and thus affects the accuracy of static features.

Distinguishing from static features, dynamic features are extracted from execution traces which make them more resilient to encryption or other obfuscation techniques. Blokhin et al. [5] partitioned system call logs acquired from sandbox into system call sequences as features. Bailey et al. [6] described malware behavior at a high level abstraction in terms of system state change profiles that the malware causes on the system, like modified registry keys, network access. Wüchner et al. [7] presented quantitative data flow graphs (QDFGs) to model program behavior through using system calls integrated with quantifiable data flow. However, many newly malware instances are able to detect the instrumented environment and refuse preforming malicious activity to evade dynamic analysis.

3 System Overview

Malware similarity analysis is divided into two steps: feature extraction and similarity calculation. Figure 1 shows the overview of our framework. A malware sample is represented as a function-call graph. Each vertex in the graph corresponds to a function and edges represent the caller-callee relationship between functions. There are two kinds of functions: local functions and external functions. For each local function, we record its execution traces with dynamic instrumentation tool Pin as its label. For external functions, we take the string of function names as the label which indicates the functionality of each function. We integrate dynamic feature into function representations as a fine-grained hybrid feature.

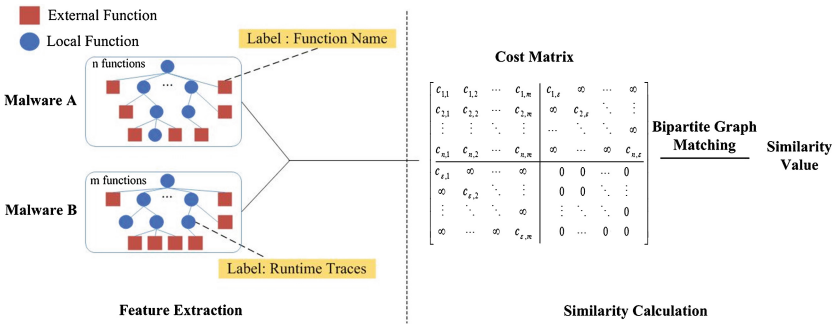


Fig. 1. The system overview

Then we use graph matching algorithm to calculate the similarity degree between pairwise samples. Graph matching methods fall into two categories: exact matching and inexact matching. Exact matching problems are NP-complete. Therefore, we use graph edit distance, one of the most widely used inexact matching algorithms. It is highly flexible and is applicable to types of graph integrated with special domain knowledge by means of cost functions. Graph edit distance is defined as the cost of the least expensive sequence of edit operations that are needed to transform one graph to another.

4 Feature Extraction

This section we first introduce the definition of function-call graph. Then, we describe the extraction of the dynamic runtime traces for each local function.

4.1 Function-Call Graph

G = (V,E) is a directed graph composed of vertex set *V* and edge set *E*. Each vertex in the graph corresponds to a function included in the program. Each edge represents the caller-callee relationship between functions. The vertex can be divided into two categories: local functions and external functions. Local functions are functions written by

malware authors. Statically-linked and dynamic-imported functions are the external functions.

Function identification is a tough challenge in binary analysis. In our paper, we use IDA Pro to identify the boundary of functions which has achieved reasonable accuracy. In IDA representation, local functions are named with "sub_xxxxx", external functions are named just the function names. IDA can provide the function-call graph of applications directly. But as pointed in [8], the performance of IDA is still imprecise, the missed function and misidentified function rate is alarming. In future, we will take this issue into account.

Each function in the graph has a label. The labels of external functions are function names. Whereas the label for local functions are the execution traces acquired from dynamic tools. We run the malicious program with instrumentation tool Pin and record each function's runtime traces to form a label vector F_l, which is composed of feature sets $F_l = \{f_1 f_2, \cdots, f_n\}$. Each f_i denotes an aspect of dynamic behavior of functions, like system calls the function invoked, the values the function written to memory, etc. We choose four feature sets: the values read from the stack f_1, the values written to stack f_2, the values read from memory f_3 and the values written to memory f_4. The feature vector is easily being extended.

5 Similarity Calculation

The central component of malware similarity analysis is to measure the distance among malware instances. The samples are represented as function-call graph, thus casts the problem into graph matching. Therefore, we use graph edit distance as a metric to weigh the similarity between variants.

Graph edit distance was first proposed in [9]. It is defined as the minimum cost amount of operation that is needed to transform one graph into another. Bipartite graph matching as an approximate computation method of graph edit distance has been proposed in [10]. It is a suboptimal method based on the procedure that mapping nodes and their local structures of one graph to nodes and structures of another graph.

Let $G_1 = (V_1, E_1, l_1)$ and $G_2 = (V_2, E_2, l_2)$ be the source and target graph as two parts of bipartite graph, where $V_1 = (v_1, \cdots, v_n)$, $V_2 = (u_1, \cdots, u_m)$. The cost matrix is as below:

$$C = \left[\begin{array}{cccc|cccc} c_{1,1} & c_{1,2} & \cdots & c_{1,m} & c_{1,\varepsilon} & \infty & \cdots & \infty \\ c_{2,1} & c_{2,2} & \cdots & c_{2,m} & \infty & c_{2,\varepsilon} & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \cdots & \ddots & \ddots & \infty \\ c_{n,1} & c_{n,2} & \cdots & c_{n,m} & \infty & \cdots & \infty & c_{n,\varepsilon} \\ \hline c_{\varepsilon,1} & \infty & \cdots & \infty & 0 & 0 & \cdots & 0 \\ \infty & c_{\varepsilon,2} & \ddots & \vdots & 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \infty & \vdots & \ddots & \ddots & 0 \\ \infty & \cdots & \infty & c_{\varepsilon,m} & 0 & \cdots & 0 & 0 \end{array}\right]$$

In our cost matrix, $c_{ij} = 1/\varepsilon_{ij}$, where ε_{ij} denotes the similarity degree of two functons. For external function nodes, ε_{ij} is measured by the longest common substring (LCS) of function names. For local function nodes, ε_{ij} is measured by Jaccard similarity coefficient between two trace vectors F_l.

6 Evaluation

To conduct the accuracy of the algorithm we proposed, we use the data set from VX Heavens, which has already been classified into malware families. The data set has 17 malware families and 1,326 samples, including Worms, Trojans, and Virus. We compared the results with [11], which used static function-call graph to compute similarity of binaries. The results demonstrate that hybrid feature outperforms solely static function-call graph.

Virus.Win32.Sality is considered as one of the most complex and formidable family of malware according to Wiki. And Table 1 shows the similarity matrix between variants in family Sality. The values of leading diagonal of the matrix are equal to 1 which demonstrates comparing with themselves. The upper triangular half of the matrix are the data of our experiment and below are [11]. The results indicate that, our method outperforms [11] when the similarity value is low within family variants. This is because variants' dynamic behaviors may keep in step with each other, although their static structures are different due to obfuscations.

Table 1. The similarity matrix of Virus.Win32.Sality

	Sality.a	Sality.c	Sality.d	Sality.e	Sality.f	Sality.g
Sality.a	1	0.991	0.941	0.618	0.612	.0867
Sality.c	0.996	1	0.933	0.618	0.612	0.862
Sality.d	0.96	0.957	1	0.651	0.639	0.889
Sality.e	0.627	0.625	0.645	1	0.961	0.856
Sality.f	0.408	0.408	0.419	0.581	1	0.848
Sality.g	0.870	0.867	0.884	0.682	0.438	1

Table 2. The similarity matrix of Email-Worm.Win32.Klez

	Klez.a	Klez.b	Klez.c	Klez.d	Klez.e	Klez.g	Klez.h	Klez.i	Klez.j
Klez.a	1	0.964	0.976	0.911	0.791	0.789	0.783	0.783	0.783
Klez.b	0.959	1	0.939	0.939	0.801	0.8	0.793	0.793	0.801
Klez.c	1	0.959	1	0.916	0.792	0.791	0.784	0.784	0.792
Klez.d	0.869	0.91	0.869	1	0.811	0.811	0.806	0.806	0.812
Klez.e	0.614	0.639	0.614	0.672	1	0.996	0.967	0.968	0.994
Klez.g	0.614	0.639	0.614	0.672	1	1	0.965	0.967	0.993
Klez.h	0.615	0.637	0.615	0.656	0.948	0.948	1	0.999	0.949
Klez.i	0.615	0.637	0.615	0.656	0.948	0.948	1	1	0.949
Klez.j	0.614	0.639	0.614	0.672	1	1	0.948	0.948	1

Table 2 shows the similarity matrix of family Email-worm.Win32.Klez. Compared to [11], our result achieves better accuracy. Although the accuracy of a few pairs of our results are little bit lower than theirs, such as pair Klez.j and Klez.g, the similarity value of their method is 1 and ours is 0.993. But for the rest of pairwise samples, our results are much higher than [11].

7 Conclusion

The large volume of malware variations poses major challenge to Anti-Virus companies. Malware similarity analysis is a critical step for malware detection and classification. In this paper, we propose a novel fine-grained hybrid feature to calculate the similarity degree between malware instances. We merge dynamic runtime traces into static function-call graph representation as a hybrid one. In evaluation, we have conducted extensive experiments with 1,326 samples in 17 families. The result shows that our algorithm is more accurate. The main contributions of our work include: (1) a hybrid graph feature for computing the similarity between malware variants to improve the accuracy of the result; (2) integrate dynamic runtime traces into function node representation in order to take advantage of both simultaneously; (3) a fully study of the performance to validate its efficiency and accuracy with a 1,326 samples database.

Acknowledgement. This work is supported by the National Science Foundation of China (No.61472439, No.61271252).

References

1. Shafiq, M.Z., Tabish, S.M., Mirza, F., Farooq, M.: PE-Miner: mining structural information to detect malicious executables in realtime. In: Kirda, E., Jha, S., Balzarotti, D. (eds.) RAID 2009. LNCS, vol. 5758, pp. 121–141. Springer, Heidelberg (2009). doi:10.1007/978-3-642-04342-0_7
2. Kolter, J.Z., Maloof, M.A.: Learning to detect and classify malicious executables in the wild. J. Mach. Learn. Res. **6**(4), 2721–2744 (2006)
3. Hu, X., Chiueh, T.-C., Shin, K.G.: Large-scale malware indexing using function-call graphs. In: Proceedings of the 16th ACM Conference on Computer and Communications Security. ACM (2009)
4. Moser, A., Kruegel, C., Kirda, E.: Limits of static analysis for malware detection **68**(6), 421–430 (2008)
5. Blokhin, K., Saxe, J., Mentis, D.: Malware similarity identification using call graph based system call subsequence features. In: IEEE International Conference on Distributed Computing Systems Workshops (2013)
6. Bailey, M., Oberheide, J., Andersen, J., Mao, Z.,Morley, Jahanian, F., Nazario, J.: Automated classification and analysis of internet malware. In: Kruegel, C., Lippmann, R., Clark, A. (eds.) RAID 2007. LNCS, vol. 4637, pp. 178–197. Springer, Heidelberg (2007). doi:10.1007/978-3-540-74320-0_10

7. Wüchner, T., Ochoa, M., Pretschner, A.: Robust and effective malware detection through quantitative data flow graph metrics. In: Almgren, M., Gulisano, V., Maggi, F. (eds.) DIMVA 2015. LNCS, vol. 9148, pp. 98–118. Springer, Heidelberg (2015). doi:10.1007/978-3-319-20550-2_6
8. Bao, T., et al.: Byteweight: learning to recognize functions in binary code. In: USENIX Security Symposium (2014)
9. Sanfeliu, A., Fu, K.S.: A distance measure between attributed relational graphs for pattern recognition. IEEE Trans. Syst. Man Cybern. **SMC-13**(3), 353–362 (1983)
10. Riesen, K., Bunke, H.: Approximate graph edit distance computation by means of bipartite graph matching. Image Vis. Comput. **27**(7), 950–959 (2009)
11. Shang, S., et al.: Detecting malware variants via function-call graph similarity. In: International Conference on Malicious and Unwanted Software (2010)

A Wireless Kinect Sensor Network System for Virtual Reality Applications

Mengxuan Li[1], Wei Song[1,4(✉)], Liang Song[2], Kaisi Huang[1], Yulong Xi[3], and Kyungeun Cho[3]

[1] Department of Digital Media Technology, North China University of Technology, Beijing, China
sw@ncut.edu.cn
[2] Beijing Innolinks Technology Co., Ltd.,, Beijing, China
[3] Department of Multimedia Engineering, Dongguk University, Seoul, Korea
[4] Beijing Key Laboratory on Integration and Analysis of Large-Scale Stream Data, North China University of Technology, Beijing, China

Abstract. Currently, Microsoft Kinect, a motion sensing input device, has been developed quickly in research for human gesture recognition. The Kinect integrating into games and Virtual Reality (VR) improves the immersion sense and natural user experience. However, the Kinect is able to accurately measure a user within five meters, while the user must face to the sensor. To solve this problem, this paper develops a wireless Kinect sensor network system to detect users at several viewports. This system utilizes multiple Kinect clients to sense user's gesture information, which is transmitted to a VR managing server for the integration of the distributed sensing datasets. Different from the VR application with a single Kinect, our proposed system is able to support the user's walking around no matter whether he is facing the sensors or not. Meanwhile, we developed a virtual boxing VR game with two Kinects, Samsung Gear VR and Unity3D environment, which verified the effective performance of the proposed system.

Keywords: Kinect · Wireless sensor network · Virtual Reality · Sensor selection

1 Introduction

Tracking the skeleton of human body from RGB image and depth sensors of the Microsoft Kinect has been widely applied for the interaction between users with virtual objects in various multimedia fields, such as Natural User Interface (NUI) and Virtual Reality (VR) [1]. The kinematics researches on the Kinect, such as human action recognition and classification, have been soaring rapidly [2]. When we utilize the Kinect to acquire user's gesture, he needs to stand in front of the Kinect with a limited distance and face to the Kinect. Otherwise, weak and inaccurate signals are sensed.

To provide an interactive environment, where users are able to walk and rotate freely, this paper proposes a wireless Kinect sensor network system for VR applications. In the system, multiple Kinect sensors installed in their respective client computers detect user's gesture information at different viewports. The sensed datasets of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_10

the distributed clients are sent to a VR managing server, which selects an adaptive Kinect based on the user's distances and orientations of the Kinect sensors. In the sensor selection process, we apply a Bivartiate Gaussian probability density function and Maximum Likelihood estimation method to choose the most suitable one to provide accurate gesture information.

By installing multiple Kinect sensors in VR user's activity environment, the proposed method extends the motion measurements of the Kinect and solves the data lost situation when the user turns back. In our method, only small datasets including user's position and body joints are transmitted between the clients and the server, which satisfied the real-time wireless network transmission requirement.

The remainder of this paper is organized as follows. Section 2 overviews related work. Section 3 describes the proposed wireless Kinect sensor network system. Section 4 develops a boxing VR game using the proposed system. Section 5 concludes this paper.

2 Related Work

Currently, Kinect is a popular display device in VR development, which reports user's localization and gesture information [3]. A single Kinect can only capture the front side of users facing to the sensor. To sense the back side, Chen et al. [4] utilized multiple Kinect to reconstruct an entire 3D mesh of the segmented foreground human voxels with color information. To track people in unconstrained environments, Sun et al. [5] proposed a pairwise skeleton matching scheme using the sensing results from multiple Kinects. Using a Kalman filter, they skeleton joints were calibrated and tracked across consecutive frames. Using this method, we found that different Kinects provided different joints localization, because the sensed surfaces were not same at different viewports.

To acquire accurate datasets from multiple sensors, Chua et al. [6] addressed a sensor selection problem in smart house using Naïve Bayes classifier, decision tree, and k-Nearest Neighbor algorithms. Sevrin et al. [7] proposed a people localization system with a multiple Kinects trajectory fusion algorithm. The system selected the best possible choice among the Kinects adaptively in order to detect people with a high accurate rate. Following these sensor selection theories, we developed a wireless and reliable sensor network for VR applications, which enable the users to walk and interact with the virtual objects freely.

3 Wireless Kinect Sensor Network System

We developed a wireless Kinect sensor network system, as shown in Fig. 1, which integrates the user's gesture datasets from multiple Kinect client into a VR server. In the multiple wireless clients, the user's skeleton motion is detected at multiple views of the Kinect sensors. The system gathers the distributed datasets via WiFi network. The best dataset is selected using a bivartiate Gaussian probability density function. In Fig. 1(b), we develop a VR application using the two Kinect sensors integration, which enables omnidirectional detection of the user.

Fig. 1. The proposed wireless Kinect sensor network system. (a) The framework of the multiple Kinect sensors integration. (b) A scene of the VR application using the proposed system with two Kinect sensors.

A Kinect is installed at each client to detect the user's gesture information at different position, which is transmitted to an application server. From several gathered datasets, effectiveness of each sensor is generated based on the user's distance d and facing orientation θ to the Kinect. If the distance is close and the orientation is facing to a sensor, the effectiveness of this sensor is high. For example, in Fig. 2, k_1, k_2, and k_3 represent the installed Kinect; d_1, d_2, and d_3 represent the distance from the person to each Kinect; θ_1, θ_2, and θ_3 represent the orientation between the user's facing direction to each Kinect. The variables d_3 and θ_3 are smallest, so that we select the Kinect k_3 to detect the user's motion.

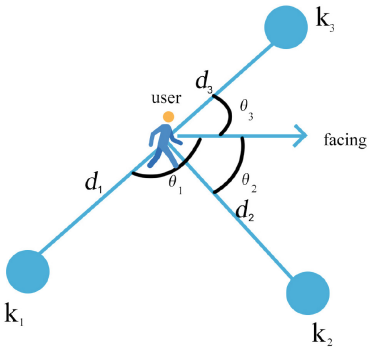


Fig. 2. An example of Kinect sensor selection.

To select the best sensor, we apply a bivartiate Gaussian probability density function (PDF) for the effectiveness estimation, formulated as follows.

$$f_{ki}(d_i, \theta_i) = \frac{\exp\left[-\frac{(\frac{d_i-d_0}{\sigma_1})^2 - \frac{2\rho(d_i-d_0)(\theta_i-\theta_0)}{\sigma_1\sigma_2} + (\frac{\theta_i-\theta_0}{\sigma_2})^2}{2(1-\rho^2)}\right]}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \quad (1)$$

Here, the variables $d \in [0 \sim \infty)$, $\theta \in [-\pi \sim \pi)$, $\sigma_1 = 1$, $\sigma_2 = 1$, $\rho \in [-1, 0]$. Through the experiments, we found that $d_0 = 5$ and $\theta_0 = 0$ is the perfect position for the Kinect detection. The best Kinect is selected by determined by a maximum likelihood function expressed as follows.

$$k = \arg\max_{k_i} f_{k_i}(d_i, \theta_i) \tag{2}$$

4 Experiments

Using the proposed system, we developed a VR boxing game shown in Fig. 3. We utilized two Microsoft Kinect2 sensors to detect user's gesture on two clients, which are 3.1 GHz Intel® Core™ i7-5557U CPU NUC mini PCs with 16 GB RAM. The VR client is implemented on a Samsung Gear VR with a Samsung galaxy Note 4 in it. The Note 4 has a 2.7 GHz Qualcomm Snapdragon Quad CPU, 3 GB RAM, 2560 * 1440 resolution, and android 4.4 operation system.

Fig. 3. A VR boxing game developed using the proposed wireless multiple Kinect sensor selection system.

In the system, the user location and orientation was detected by the two Kinects. When the player was facing to a Kinect with a distance between 2 to 6 meters, the motion information was sensed precisely. By selection of effective Kinect, the user was able to take free movement and interact with the virtual boxer at omnidirectional orientation. Meanwhile, the monitor of the server rendered the game visualization result synchronously with the VR display. The processing speed of our application including data sensing, transmission, and visualization was more than 35 fps, which achieved a real-time approach.

5 Conclusions

To provide a free movement environment for VR application, this paper demonstrated an effective wireless Kinect sensors selection system via the WiFi network. Using the bivartiate Gaussian PDF, we estimated the effectiveness of each Kinect sensor based on the user's distance and orientation to it. A maximum likelihood function was applied to select the best sensor for providing accurate dataset of user's motion. Using the proposed system, we developed a VR boxing game with two Kinects, where the user was able to move and interact with virtual objects freely. In future, we will integrate some touching sensors to the system for user experience enhancement.

Acknowledgment. This research was supported by the National Natural Science Foundation of China (61503005), by the research foundation of NCUT (XN070027, XN001-93, and XN001-83), by Advantageous Subject Cultivation Fund of NCUT, and by SRF for ROCS, SEM.

References

1. Ales, P., Oldrich, V., Martin, V., et al.: Use of the image and depth sensors of the Microsoft Kinect for the detection of gait disorders. Neural Comput. Appl. **26**, 1621–1629 (2015)
2. Mohammed, A., Ahmed, S.: Kinect-Based Humanoid Robotic Manipulator for Human Upper Limbs Movements Tracking. Intell. Control Autom. **6**, 29–37 (2015)
3. Chen, H., Dai, Z., Liu, Z., et al.: An image-to-class dynamic time warping approach for both 3D static and trajectory hand gesture recognition. Pattern Recogn. **55**, 137–147 (2016)
4. Chen, Y., Dang, G., Chen, Z., et al.: Fast capture of personalized avatar using two Kinects. J. Manufact. Syst. **33**, 233–240 (2014)
5. Sun, S., Kuo, C., Chang, P.: People tracking in an environment with multiple depth cameras: a skeleton-based pairwise trajectory matching scheme. J. Vis. Commun. Image Represent. **35**, 36–54 (2016)
6. Chua, S.L., Foo, L.K.: Sensor selection in smart homes. Procedia Comput. Sci. **69**, 116–124 (2015)
7. Sevrin, L., Noury, N., Abouchi, N., et al.: Preliminary results on algorithms for multi-kinect trajectory fusion in a living lab. IRBM **36**, 361–366 (2015)

Finding Comfortable Settings of Snake Game Using Game Refinement Measurement

Anunpattana Punyawee[1], Chetprayoon Panumate[2(✉)], and Hiroyuki Iida[2]

[1] School of Information, Computer and Communication Techonolgy, Sirindhorn International Institute of Technology, Thammasat University, Bangkok, Thailand
punyawee.anunpattana@gmail.com
[2] Research Center for Entertainment Science, Japan Advanced Institute of Science and Technology, Nomi, Japan
{panumate.c, iida}@jaist.ac.jp

Abstract. This paper explores the attractiveness and sophistication of Snake game which originated as an arcade maze game and has been very popular for the decades on mobile phones platform. It presents an approach to find comfortable settings of Snake game by using game refinement theory. Basic AI is created for collecting the data instead of human in this research. The results obtained show the reason why Snake game has been so popular on mobile phones and people can feel entertaining and excited.

Keywords: Game refinement theory · Snake game · Arcade game

1 Introduction

Snake game is [1] a type of arcade maze games originally developing from "Blockade" which has been developed by Gremlin Industries and published by Sega in October 1976 [6]. There are many clone games based on Blockade game inspiration, e.g., Bigfoot Bonkers, Surround, Dominos, etc. [2, 4]. Snake games are considered to be a skillful game, players try to achieve maximum score as high as possible. We show, in Fig. 1 a screenshot of Snake game on Nokia 3310.

Nokia is well-known for putting Snake game in their phones [5, 13]. Towards the end of the year 2000, Nokia released one of the most successful phones, Nokia 3310, whose Snake game is so popular. Snake games are still included in some new phones from Nokia and available for all platforms. The history of Snake games on Nokia mobile phones [14] is shown in Table 1.

We can control the snake in Snake game by using the four direction buttons relative to the direction it is heading in. The snake increases its speed as it gets longer by eating

Snake Game® is a registered trademark. All intellectual property rights in and to the game are owned by its respected owner.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_11

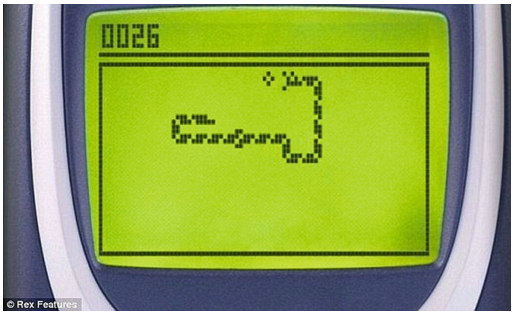


Fig. 1. A screenshot of Snake game on Nokia 3310

Table 1. A historical overview of Snake games on Nokia mobile phones

Year	Version
1997	Snake
2000	Snake II
2002	Snake EX
2003	Snake EX2
2005	Snakes
2005	Snake III
2006	Snake Xenzia
2008	Snakes Subsonic

fruits. The goal of this game is to collect fruits as many as possible. A player loses when the snake crashes the wall or crashes itself [3].

However, some variants have the regulation that the snake's speed will go up when eating fruits and the snake can pass through the wall, then appear on the opposite side. Actually, there are many factors affected to Snake game. It can be summarized as shown below.

Notation 1. *A family of snake games can be denoted by SNG(s, a, m, c, l) with the initial physical speed of snake in frame per second s, acceleration of snake when eating each fruit in frame per second a, size of the map m specifying in terms of total area (pixel* $\times$ *pixel), wall condition c specifying trigger between 0 and 1 which means no wall and have a wall respectively, and initial length of snake in pixel l.*

While there are many variants of Snake game, some of them are so popular and some of them are not. The main reason behind this fact is the setting of some variables in the game. Its setting may often considerably affect the entertaining aspect which leads the game to be exciting or boring. That is the reason why we assume that each setting of video game has its own reasons behind this, and therefore finding a comfortable setting can be significantly important to maximize the entertainment in playing the game considered. Thus, this paper presents an approach to find comfortable settings of Snake game. To tackle this challenge, game refinement theory [11] is used as an

important tool for the assessment. We consider mathematical equations and a reasonable game information progress model to derive a game refinement measure for Snake game and its variants. Then *SNG(s, a, m, c, l)* are considered and evaluated by adjusting each factor.

This paper is organized as follows. Section 2 provides the fundamental ideas of game refinement theory, including our approach to the application of Snake game. Experimental results are shown and discussed in Sect. 3, which we implement the simple game to collect the data based on AI. Finally, Sect. 4 gives concluding remarks and considerations for further works.

2 Assessment Methodology

A general model of game refinement was proposed based on the concept of game progress and game information progress [7]. It bridges a gap between board games and sports games. Game information progress presents the degree of certainty of game's results in time or in steps. Having full information of the game progress, i.e., after its conclusion, game progress $x(t)$ will be given as a linear function of time t with $0 \leq t \leq t_k$ and $0 \leq x(t) \leq x(t_k)$, as shown in Eq. (1).

$$x(t) = \frac{x(t_k)}{t_k} t \tag{1}$$

However, the game information progress given by Eq. (1) is unknown during the in-game period. The presence of uncertainty during the game, often until the final moments of a game, reasonably renders game progress as exponential. Hence, a realistic model of game information progress is given by Eq. (2).

$$x(t) = x(t_k)\left(\frac{t}{t_k}\right)^n \tag{2}$$

Here n stands for a constant parameter which is given based on the perspective of an observer of the game considered. Then acceleration of game information progress is obtained by deriving Eq. (2) twice. Solving it at $t = t_k$, we have Eq. (3).

$$x^{''(t_k)} = \frac{x(t_k)}{(t_k)^n} t^{n-2} n(n-1) = \frac{x(t_k)}{(t_k)^2} n(n-1) \tag{3}$$

It is assumed in the current model that game information progress in any type of game is encoded and transported in our brains. We do not yet know about the physics of information in the brain, but it is likely that the acceleration of information progress is subject to the forces and laws of physics. Therefore, we expect that the larger the value $\frac{x(t_k)}{(t_k)^2}$ is, the more the game becomes exciting, due in p part to the uncertainty of game outcome. Thus, we use its root square, $\frac{\sqrt{x(t_k)}}{t_k}$, as a game refinement measure called R value.

The previous studies apply game refinement theory to many domains of game such as board game [8], video games [10] and sports [9] by figuring out a suitable model for the game under consideration. It has been confirmed that the sophisticated games will have R value in the sophisticated zone, 0.07–0.08. The results from the previous studies are shown in Table 2.

Table 2. Measures of game refinement for board games and sports

Game	R
Chess	0.074
Go	0.076
Basketball	0.073
Soccer	0.073

Table 3. A historical overview of Snake games on Nokia mobile phones

	c	G	T	R
Human	0	16.5	72.6	0.056
	1	6.1	31	0.08
AI	0	4.44	34.98	0.06
	1	2.18	17.03	0.087

In soccer and basketball [12], a game progress model is constructed with a focus on the number of goals G and the number of attacks or shot attempts T. We then obtain $R = \frac{\sqrt{G}}{T}$. With the same idea, in Snake game, we construct a functional game progress model with a focus on the average total scores per game (say G) and average number of moves counted from the number of clicking button (say T) as shown in Eq. (4).

$$R = \frac{\sqrt{G}}{T} = \frac{\sqrt{average\ total\ scores\ per\ game}}{average\ total\ number\ of\ moves\ per\ game} \tag{4}$$

3 Analyzing Snake Game Variants

We demonstrate an analysis of Snake game variants *SNG(s, a, m, c, l)* for finding comfortable settings. For this purpose we develop a computer player (AI) of playing Snake game.

3.1 Developing Snake AI

We first develop the AI for snake game in order to perform many experiments. The basic objective of ordinary players is to eat fruits with low move and avoid hitting itself. However, most human players would have their own mistakes, so AI should be randomized in the position of snake head for miss turning of snake to the fruits. Due to the limitation that AI cannot make a decision spontaneously like a human, we try to implement AI which can be assumed as a human player as much as possible, which is described in Algorithm 1. Then, Table 3 shows the comparison of results using AI and human. For each setting, we play 250 games and the version used for this test is the replica of the game as it currently exists on Nokia. The value of G and T which are collected from AI might be slightly different from human data because human players

can develop their skills while playing. However, the *R* value is not significantly different, so we can use AI instead of human. Moreover, we will see that, with the wall condition, game will be more difficult because it is easier to die, so it makes *R* value higher

Algorithm 1 BasicPlayerAI

```
1: procedure DirectionDecision
2:     i ← random from 0 to 2
3:     fruitPos ← random fruit position
4:     snakeX and snakeY ← position of snake's head
5:     headIndex ← i
6:     snakeX and snakeY ← headIndex
7:     if snakeY > fruitPos then
8:         direction = up
9:     else if snakeY < fruitPos then
10:        direction = down
11:    else if snakeY = fruitPos then
12:        if snakeX > fruitPos then
13:            direction = left
14:        else if snakeX < fruitPos then
15:            direction = right
16:    end if
17:    return direction
```

3.2 Analysis of *SNG(s, 1, 240, 1, 3)*: Speed

Table 4 shows the measures of game refinement for the analysis of snake's speed with $5 \leq s \leq 30$. If the speed is high, the controlling is too difficult which will increase the number of moves for surviving. Increasing *s* means increasing *T* because player will make many mistakes while moving with the fast speed, so the number of moves, *T*, is increased and it leads *R* value decreased respectively.

Table 4. Measures of game refinement for *SNG(s, 1, 240, 1, 3)*

s	G	T	R
5	5.51	12.8	0.183
10	4.28	11.8	0.175
15	2.18	17.03	0.087
20	2.54	19.46	0.082
25	2.34	20.07	0.076
30	2.06	20.24	0.071

Table 5. Measures of game refinement for *SNG(15, a, 240, 1, 3)*

a	G	T	R
0.2	2.26	16.85	0.089
0.5	2.31	18.62	0.082
1	2.18	17.03	0.087
2	1.96	16.024	0.087
3	2.43	18.88	0.083
4	2.32	21.92	0.079

3.3 Analysis of *SNG(15, a, 240, 1, 3)*: Acceleration

Table 5 shows the measures of game refinement for snake's acceleration, a. In detail, the information about acceleration cannot clearly observe while playing because players should focus on their own turn in Snake game. In game progress and case of higher speed, players will die quickly if they reach at some points. In sense of entertaining, acceleration is quite less impact to this game progress model. When game is progressing for many games, the average number of moves is the same on each value of acceleration.

3.4 Analysis of *SNG(s, 1, m, 1, 3)*: Speed and Map Size

Table 6 shows the measures of game refinement for snake's speed in different map size. We focus on the values of s and m with $15 \leq s \leq 30$ and m = 160, 240, 336, 448, 960 *pixel x pixel*. The size of map will control the game length because if the map is larger, it will increase the number of survival ways for players. By adjusting these factors, we can identify the comfortable settings of Snake game.

Table 6. Measures of game refinement for *SNG(s, 1, m, 1, 3)*

s	m	G	T	R
15	160	1.80	12.8	0.112
20	160	1.96	13.7	0.102
25	160	1.91	14.23	0.097
30	160	1.96	14.50	0.0.96
15	240	2.18	17.03	0.087
15	336	2.50	20.33	0.079
20	336	2.30	19.92	0.076
25	336	2.40	20.90	0.074
30	336	2.52	21.67	0.073
15	448	2.44	22.40	0.070
20	448	2.62	22.43	0.072
15	960	3.12	27.80	0.063

Table 7. Measures of game refinement for *SNG(15, a, 240, 1, 3)*

s	l	G	T	R
5	3	5.51	12.80	0.183
15	3	2.18	17.03	0.087
30	3	2.06	20.24	0.071
5	6	4.79	11.09	0.197
15	6	1.25	10.18	0.110
30	6	1.16	10.23	0.105
5	9	4.55	10.48	0.200
15	9	0.92	17.40	0.130
30	9	0.82	8.90	0.102

3.5 Analysis of *SNG(s, 1, 240, 1, l)*: Speed and Length of Snake

Table 7 shows the measures of game refinement with a focus on the initial snake's length and speed. Each variant of Snake game has different length of snake and speed. In practical, snake length is difficult to see clearly in player's view because the screen does not show obviously pixel. However, a big gap in different length of snake will affect to the difficulty obviously. By adjusting these factors, we can find the snake game with the comfortable settings, $R = 0.07–0.08$, by setting *SNG(30, 1, 240, 1, 3)*.

4 Concluding Remarks

Based on the experiments, we analyzed what factors had more effect to the entertainment of the game. We explored that the physical speed of snake, map size and wall condition can be seen impact apparently in player's view. Under the assumption that the sophisticated zone of R value is around 0.07–0.08 which many previous studies confirm, we found the comfortable settings of Snake game for *SNG(s, a, m, c, l)*. First, s is around 25–30 for *SNG(s, 1, 240, 1, 3)*. Secondly, *SNG(15, 1, 240, c, 3)* can play either $c = 0$ and $c = 1$, condition of the wall indicates the difficulty of the game. In the same manner with *SNG(15, 1, m, 1, 3)*, it is best to play with the appropriate map size as same as a mobile phone. For other analysis, we explored that acceleration a and length of snake l are less impact to game entertainment, this is because these two variants are subsections of speed. Further works can improve in many points such as improving the quality of AI, adding more factors and investigating in another domain.

References

1. Chittaro, L., Sioni, R.: Turning the classic snake mobile game into a location–based exergame that encourages walking. In: Bang, M., Ragnemalm, Eva, L. (eds.) PERSUASIVE 2012. LNCS, vol. 7284, pp. 43–54. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31037-9_4
2. DeMaria, R., Wilson, J.: High Score!: The Illustrated History of Electronic Games. Computer Games Series. McGraw-Hill/Osborne (2004). https://books.google.co.jp/books?id=HJNvZLvpCEQC
3. Ehlis, T.: Application of genetic programming to the "snake game". Gamedev. Net 175 (2000)
4. Goggin, G.: Global Mobile Media. Media and Communication Studies. Routledge, New York (2011). https://books.google.co.jp/books?id=UbTVmAEACAAJ
5. Haikaio, M., et al.: Nokia-the inside story (2002)
6. Arcade history: An arcade game history site (2016). http://www.arcade-history.com/
7. Iida, H., Takahara, K., Nagashima, J., Kajihara, Y., Hashimoto, T.: An application of game-refinement theory to Mah Jong. In: Rauterberg, M. (ed.) ICEC 2004. LNCS, vol. 3166, pp. 333–338. Springer, Heidelberg (2004). doi:10.1007/978-3-540-28643-1_41
8. Iida, H., Takeshita, N., Yoshimura, J.: A metric for entertainment of boardgames: its implication for evolution of chess variants. In: Nakatsu, R., Hoshino, J. (eds.). ITIFIP, vol. 112, pp. 65–72. Springer, Heidelberg (2003). doi:10.1007/978-0-387-35660-0_8
9. Nossal, N., Iida, H.: Game refinement theory and its application to score limit games. In: 2014 IEEE Games Media Entertainment (GEM), pp. 1–3. IEEE (2014)
10. Panumate, C., Xiong, S., Iida, H.: An approach to quantifying pokemon's entertainment impact with focus on battle. In: 2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence (ACIT-CSI), pp. 60–66. IEEE (2015)
11. Sutiono, A.P., Purwarianti, A., Iida, H.: A mathematical model of game refinement. In: Reidsma, D., Choi, I., Bargar, R. (eds.) INTETAIN 2014. LNICSSITE, vol. 136, pp. 148–151. Springer, Heidelberg (2014). doi:10.1007/978-3-319-08189-2_22

12. Sutiono, A.P., Ramadan, R., Jarukasetporn, P., Takeuchi, J., Purwarianti, A., Iida, H.: A Mathematical Model of Game Refinement and Its Applications to Sports Games (2015)
13. Team, M.D.: The evolution of snake. http://blogs.windows.com/devices/2010/11/25/the-evolution-of-snake/ (2016)
14. Wikipedia: Wikipedia snake game page (2016). https://en.wikipedia.org/wiki/Snake. (video game)

Code Modification and Obfuscation Detection Test Using Malicious Script Distributing Website Inspection Technology

Seong-Min Park(✉), Han-Chul Bae, Young-Tae Cha, and Hwan-Kuk Kim

Security Industry Technology Division, Korea Internet & Security Agency, Seoul, Republic of Korea
{smpark, hcbae, ytcha, rinyfeel}@kisa.or.kr

Abstract. The use of non-standard plug-ins has been unavoidable in existing HTML. Accordingly, the rising dependence of web content on plug-ins increased and need for additional development suited to platform incurred considerable developmental expenses and time. HTML5 was a suggested solution to this problem, but could cause new security threats through newly added technologies, meaning that web-related elements such as websites, web content, devices and others are exposed to a new type of threat. In this thesis, we suggest a technology to detect website security threats in advance that includes HTML5.

Keywords: HTML5 · Script-based cyber attack · Web content security · Web scanner · Web content analysis

1 Introduction

According to Microsoft, attempted attacks using HTML and JavaScript vulnerability are rapidly increasing. In particular, methods such as web page built-in JavaScript functions, code obfuscation, external JavaScript calling and others have made recent web attacks possible without need for malicious code distribution and infection. Thus detection can be a difficult task with packet-based detection through security devices such as existing web firewalls, IPS and others. Besides, vulnerability to attacks goes up through newly added API and tag after HTML5 standard finalization, and tracking can be difficult since no traces like malicious code are left. And technology to inspect HTML5 security vulnerability is not yet available worldwide [1].

In this situation, we studied a technology to inspect websites that distribute malicious script using HTML5 and analyzed a result tested using such fact. In this thesis, only the core part of the details is summarized, and its composition is as follows. Section 2 examines preceding research on website inspection technology, and Sect. 3 introduces the website inspection technology we studied. Sections 4 and 5 contain our conclusions based on the verification results of the technology we studied.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_12

2 Related Work

2.1 Security Threat of HTML5

In HTML5, web attacks using new API, tag, and others became possible. Black Hat USA 2012 released HTML5 security vulnerability TOP 10, and researches on HTML5 vulnerability have been released continuously since then [2].

2.2 Web Contents Analyzing Technology

In terms of web contents analysis technology, SpyProxy and WebShield collect traffics in the network level and execute malicious scripts directly, through virtual environment or sandbox based modified virtual browser, and suggest a technology that transmits only malicious scripts eliminated safe contents to client [3].

However, as SpyProxy that collects website information using open source Squid cannot be implemented to virtual environment browser as actual user environment, it makes perfect verification difficult. In addition, for user to use web service, there is a problem of waiting time period until safe contents verification is complete. To resolve such problem, we improved performance using contents split verification technology, but it is quite inconvenient to install additional Agent on host [4].

To resolve problem of behavior monitoring based SpyProxy, WebShield suggests middle box framework creating DOM data structure instead of client browser. Client side Agent is not required. Instead, as webpage data is retained in memory, real time processing becomes possible. However, as result of test, when proxy sandbox occupies 100Mbytes memory, only 82 people per second at the maximum were allowed to have access. In other words, it means there is actual burden that minimum 10 machines are required in enterprise environment. Furthermore, it contains problem of performance deterioration due to User Script handling and frequent DOM (Document Object Model) structure updates, and with the limit to detecting HTML5 based malicious scripts [5].

3 Inspection Technology for Malicious Scripts Distributed Website

As discussed in Sect. 2, because of the performance problem of real time web contents analyzing technology, and limits of prior website inspection tools, new type of frame is necessary to cope with web based attacks. Accordingly, we intend to implement a technology that inspects vulnerabilities by visiting websites before clients use actual services, in the form of a framework.

This structure is divided into website data crawling collection technology, and analyzing technology that inspects collected data. In addition, analyzing technology consists of scripts manipulation inspection, which inspects code modification and obfuscation inspection that extracts original scripts from code obfuscation scripts.

3.1 Website Contents Collection Technology

First step to analyze and prevent website vulnerability in advance, is to collect website contents, an analysis object. Collection technology consists of crawler that performs web contents crawling, and Agent that delivers collection targets, monitors and manages crawler status, and crawler manager that assigns tasks to crawlers, saves collection result into DB.

3.2 Website Vulnerability Analyzing Technology

Inspection is performed for collected website data to figure out if there are vulnerabilities through website vulnerability analysis technology. Largely, 2 types of inspections are carried out which include scripts obfuscation inspection and modification inspection.

Attacker launches attack using scripts obfuscation for malicious website scripts not to be detected [6]. Therefore, obfuscation script should be identified and de-obfuscation needs to be carried out. First, to inspect obfuscation status of script code, measure Entropy element, N-Gram Entropy element, Max Word Size element, and then calculate obfuscation score [7].

Entropy element is to measure distribution status of bytes within web documents, and expression is as follows (b = count of each byte, T = total count of bytes, N = count of strings)

$$E(B) = \sum_{i=1}^{N} \left(\frac{b_i}{T}\right) \log\left(\frac{b_i}{T}\right) \begin{cases} B = b, i = 0, 1, \ldots, N \\ T = \sum_{i=1}^{N} b_i \end{cases} \tag{1}$$

N-Gram Entropy element measures distribution status of special characters within web documents, and then the ratio to entire Entropy.

$$R(S) = \frac{E(S)}{E(B)} * 100 \tag{2}$$

Max Word Size element measures length of the longest character string within web documents, and if each measured value exceeds given tolerance, final obfuscation score is calculated by adding up score of each element.

Obfuscation code was decoded by executing obfuscation scripts classified in such way, with V8 JavaScript engine. And we determined malicious behavior by extracting web page scripts calling data through API hooking.

Second, script modification inspection verifies if scripts are manipulated to cause malicious behaviors. If collected web contents are HTML based documents, DOM data is generated and scripts extracted. Hash value from extracted DOM data or JS file is extracted and saved. Non-cryptographic hash algorithm such as MurmurHash, CRC32, SuperFastHash, and others is used for hash value extraction. In general, while web page contents change frequently, scripts do not change unless corresponding website is

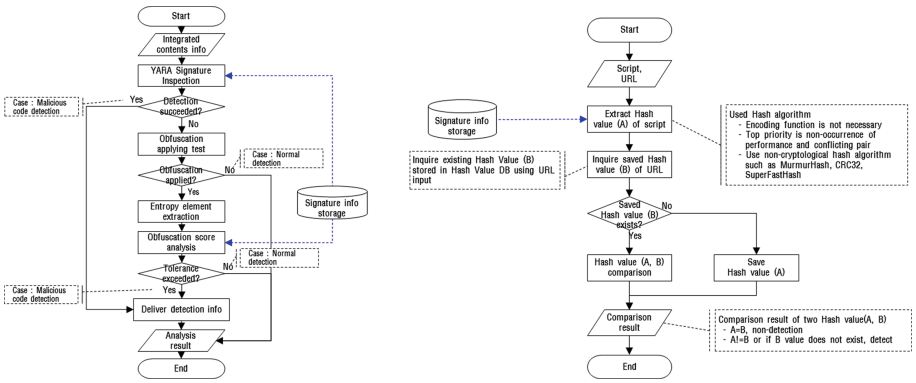


Fig. 1. Script obfuscation inspection algorithm and Scripts modification inspecting algorithm

restructured. Therefore, once hash values are saved in URL unit, based on the value as a reference, it can be compared with the hash value extracted when next user visits the same URL. As result of comparison, if hash value is the same, modification did not occur, and if different, it is regarded as modification, and static analysis is conducted (Fig. 1).

4 Code Modification and Obfuscation Detection Test

We tested the inspection technology for malicious scripts distributing website, which is implemented using 17 types of attack samples. Samples are made up to implement attack behaviors such as Scanning, Hijacking, DDoS triggering, data intercepting, and others, by injecting malicious scripts in websites. After injecting each malicious script into websites implemented for the test, we executed one test at a time. When accessing test websites from tester, it should go through inspection system that developed technology is applied, in order to verify detection process. In addition, in terms of each malicious script type, we tested total 4 types of various attack mechanisms by applying 2 types of code modification methods and 2 types of obfuscation methods (Table 1).

As result of test, 100 % detection rate was shown for original code of each attack script, but for codes that applied variable name or function name during code modification, 58.8 % of low detection rate was shown. However, with 100 % detection rate for code modified by applying Code Split method and obfuscation code applied with Packed encode method, we could confirm that it is detected being dependent upon signature ID.

Secondly, we conducted test on the performance of implemented inspection technology with commercial website as subject. Designated as safe websites by Google Safe Browsing and Norton Safe Web among Alexa Top 100 websites, 50 websites were selected. In addition, test was conducted using 50 safe websites such as Amazon, E-bay, Alibaba, and others, and 10 websites converted to HTML5 operating in Korea as subject [8].

Number of URLs that can be collected as test measurement items, and number of misdetections were used as subject. Depth5 level crawling was performed by inspection

Table 1. 17 types of malicious script samples used in test.

Sample No.	Attack	Description
1	Network Scan	Without user knowing, confirm user's internal network or port open status by using XHR(XML Http Request), WebSocket, WebRTC
2	Port Scan	
3	Cross Site WebSocket Hijacking	Without user knowing, execute WebSocket communications and establish connection with WebSocket server
4	ClickJacking (Drag&Drop)	Induce user's click or user's file (using Drag-Drop API, File API) to other places by overlapping transparent frame set as opacity (transparency) property 0, with other frame
5	ClickJacking (Click info)	
6	WebSocket data intercepting	Intercept WebSocket communication details by redefining onmessage event handler of WebSocket
7	SVG Keylogger	Send request to attacks to attacker based on user's key input through accessKey event of SVG(Scalable Vector Graphics), a XML type image
8	MouseLogger	Obtain touch event coordinates of user mouse using Pointer API
9	Cross Site Printing	Induce to print GET message by sending it to basic port 9100 of printer using XHR, WebSocket
10	Cookie Sniffing	Without user knowing, transmit the cookies within browser to attacker
11	Geolocation	Figure out device location using Geolocation API and send to attacker
12	Server-Sent Event Bot	Insert to scripts and modify web pages using Server-Sent Event, the API that provides Polling function
13	Worker DDoS	Without user knowing, trigger DDoS using Web Worker that provides separate JavaScript threads function in web page
14	WebStorage Leak	Inquire Web Storage information and transmit externally
15	IndexedDB Leak	Intercept WebSQL details which is managed by each domain within browser (However, WebSQL was eliminated from standard, and instead, Indexed DB is selected as standard)
16	Vibration Attack	Trigger user inconveniences by causing continuous vibrations in web pages
17	Script DoS (for statement)	Send request externally whenever web page connection occurs by putting scripts that send request externally into web pages

subject website, and we were able to measure number of URLs collected per second. Number of entire URL collection was 33,698, with 15,579 s required. Number of URLs collected per second based on such fact was 2.16 EA, and URL collection capacity of 1 crawler per day was measured as total 186,887 EA (Table 2).

Table 2. Performance index.

Item	Result
Number of total URL collected	33,698
Total time spent (sec)	15,579
Number of URL collected per (URL/s)	2.16
Number of URLs collected per day (1 crawler)	186,887

In addition, we conducted analysis on misdetection results along with detection results on subject sites. Number of total URLs, attempted in the analysis was 10,464 EA, 87 cases of them were confirmed as misdetection. Most of detection items confirmed as misdetection were Clickjacking attack by signature Rule ID No. 11. Clickjacking is a type of attack that sets opacity property in scripts to 0, and induces user click to malicious scripts by overlapping transparent frame with other frame. As result of applying sample HTML document about Clickjacking to YARA tool, we were able to confirm that most of detections were due to script jquery.js. Opacity property is set when suing jquery.js, and thus misdetection was occurred.

5 Conclusion and Future Work

This thesis suggests a technology to collect and analyze contents such as HTML, SCRIPT, CSS, and others, in order to conduct inspection of websites distributing malicious scripts including HTML5. Inspection test including attack samples detection, and commercial websites was conducted. In terms of performance, this technology is differentiated from existing real time malicious scripts detection technology such as SpyProxy and WebShield. In enterprise environment, when intending to collect and analyze one million web contents per day, WebShield requires 10 machines at the minimum, but with this technology, far less number of machines are needed to achieve the same goal. As number of URLs collected by 1 crawler is 186,887 EA, only 6 (5.35) machines is required to collect 1 million URLs per day.

However, some improvement needs to be made for part of test process. In commercial network website test, dynamically generated web pages were not collected during collection process using web crawling tools, instead of direct visit. Furthermore, collection was not possible due to IP blocking by web service providing company. To resolve such problem, by granting 1 s interval for each collection depth, number of URLs collected per second was reduced more than expected.

Acknowledgments. This study was conducted as part of information communications and broadcasting R&D project by Ministry of Science, ICT and Future Planning, and Institute for Information & Communications Technology Promotion [B0101-15-0230, Prevention of script based cyber attack and development of counter technology].

References

1. Microsoft Security Intelligence Report (SIR) Volume 17, pp. 43–54. Microsoft (2014)
2. Shah, S.: Founder & Director, Blueinfy Solutions, "HTML5 Top 10 Threats Stealth Attacks and Silent Exploits", pp. 1–20. Black Hat, USA (2012)
3. Agten, P., Van Acker, S., Brondsema, Y., Phung, P.H., Desmet, L., Piessens, F.: Jsand: complete client-side sandboxing of third-party javascript without browser modifications. In: Proceedings of the 28th Annual Computer Security Applications Conference, pp. 11–21. ACM (2012)
4. Moshchuk, A., et al.: SpyProxy: Execution- based Detection of Malicious Web Content. USENIX Security (2007)
5. Li, Z., et al.: WebShield: enabling various web defense techniques without client side modifications. In: NDSS (2011)
6. Xu, W., Zhang, F., Zhu, S.: The power of obfuscation techniques in malicious JavaScript code: A measurement study. In: 7th International Conference on Malicious and Unwanted Software (MALWARE). IEEE (2012)
7. Choi, Y., Kim, T., Choi, S., Lee, C.: Automatic detection for javascript obfuscation attacks in web pages through string pattern analysis. In: Lee, Y.-h., Kim, T.-h., Fang, W.-c., Ślęzak, D. (eds.) FGIT 2009. LNCS, vol. 5899, pp. 160–172. Springer, Heidelberg (2009). doi:10.1007/978-3-642-10509-8_19
8. Alexa Top 100. http://www.alexa.com/topsites

Initialization of Software Defined Wireless Bacteria-Inspired Network Platform

Shih-Yun Huang[1], Hsin-Hung Cho[2(✉)], Yu-Zen Wang[3], Timothy K. Shih[2], and Han-Chieh Chao[1,3]

[1] Department of Electrical Engineering, National Dong Hwa University, Hualien, Taiwan, R.O.C.
deantt67@gmail.com
[2] Department of Computer Science and Information Engineering, National Central University, Taoyuan, Taiwan, R.O.C.
awp.boom@gmail.com, timothykshih@gmail.com
[3] Department of Computer Science and Information Engineering, National Ilan University, Yilan, Taiwan, R.O.C.
wuc623@gmail.com, hcc@niu.edu.tw

Abstract. Fifth generation mobile network is a development trend for information and communication industry. In order to meet requirement which presented by some organizations which are working on 5 G standards and regulations, we had proposed a Software-Defined Wireless Bacteria-Inspired Network (SDWBIN) platform. In this paper, we will introduce how we initialize this prospective platform.

Keywords: 5 G mobile communication · Software-define networking · Bacteria-inspired method

1 Introduction

The information and communications industry are the development priorities for every countries, because every people think that almost all of the jobs can be completed through internet technology. The most importance is that these services can be used anytime and anywhere as long as the user's device has license for accessing of mobile communication. Such habits have made mobile communication becomes one of promising technology, hence mobile communication technology had been rapidly evoluted from 1 G to 4 G in just 30 years [1].

However, although 4 G provides 1 Gbps data rate for the static network and 100 Mbps data rate for high mobility network, it has already no potentiality to carry the explosively grown network traffic and user's expected quality of service (QoS). In order to meet demands of future information and communications industry, many organizations has begun to develop 5 G standard such as International Telecommunication Union (ITU), European Telecommunications Standards Institute (ETSI) and so on. No matter which organization agreed that next generation mobile communication must achieve 10 to 100 times of 4 G network capacity and so on. These requirements will be the major challenge for 5 G deployment. For the blueprint of future 5 G, Mobile

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_13

and Wireless Communications Enablers for the Twenty-twenty Information Society (METIS) consider that 5 G not only retain the existing technologies but also extend several new technologies which include device-to-device (D2D) communication, massive machine communication (MMC), ultra-dense networks (UDN) and so on. Hence we can know that future 5 G must be a heterogeneous network (HetNet) [2] so that the whole network environment will be difficult to manage. It means that this problem cannot be improved just by hardware solution [3]. In the previous work, we had present an idea about the Software Defined Wireless Bacteria-Inspired Network (SDWBIN) platform [4]. It is a software solution for easy management of 5 G. It means that this method is able to configure easily and not spend extra cost. In this paper, we will explain how to use bacteria behavior to deliver network traffic as well as introduce how to implement this platform.

The rest of the paper is organized as follows. Section 2 introduces background of Software Defined Wireless Bacteria-Inspired Network. Section 3 will give a clear initialization process of this platform. Finally, we will summarize research contributions and discussing the future works.

2 Background and Related Works

2.1 Software Defined Wireless Bacteria-Inspired Network

In order to carry huge network traffic, we try to let the network process to imitate the behavior of bacteria [5]. The reason is that we found the ecology of bacteria is a stable, efficient as well as this ecology has very high ability to repair itself. In general, bacteria have three main steps which are perception, infection and diffusion respectively. The perception means that each bacteria is able to sense changes in the environment and status of neighbor bacteria, hence bacteria always can find a host which has a suitable living environment to support basic conditions for survival of bacteria. However, hosts may die so that bacteria must continue to expand their living space so that they have to "infect" other hosts. In the some cases, bacteria cannot success to expand, because the such bacteria suitable species may already be endangered as well as others species cannot support the same living environment with original specie, hence bacteria must be mutated so that they have ability to suit for new environment or to infect the other species such as zoonotic diseases (Bird Flu: H5N1). For better efficiency of infection, bacteria will produce a large number of copies to sound out their neighbors and further infect them. This behavior is so called "proliferation".

However, the mentioned phenomenons belong to biological behavior. They cannot directly apply to the engineering so that we can only mimic the strong point of biological characteristics. In other words, we still keep carries of original mobile communication such as radio frequency by 3 G–5 G, UMTS instead of the water, air or any biological elements. Therefore, we cannot call our method a bacteria network, but a bacteria-inspired network. In order to successfully complete this innovative approach, we must map the bacteria phenomenon to the engineering firstly that can be referred by [4].

By previous mentions, we know that "perception" is an initial step in the bacteria ecology. In order to achieve it, we will implement the basic network platform then

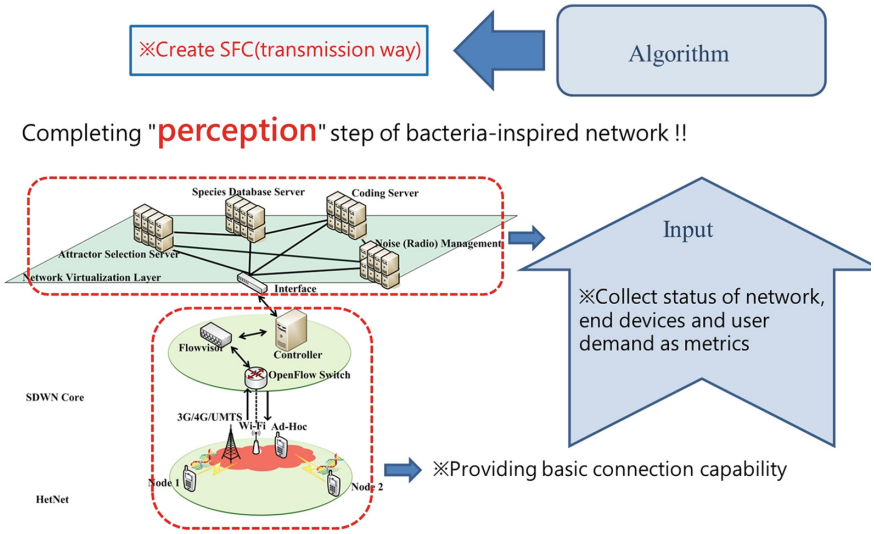


Fig. 1. Correspondence table

collects status of network, end devices and user demand as metrics. Finally, these collections will input to a policy algorithm then create a robust path to carry the network traffic so that this method will be like bacteria to perceive and find a favorable host that is shown in Fig. 1. In this paper, we will clearly introduce how to initialize our SDWBIN platform.

3 Implementation

The main concept of perception is that bacteria can sense the surrounding environment so that any bacteria are able to find a favorable host to parasitize. It means that to implement a bacteria-inspired network platform, a whole view is necessary. Therefore, we think that software-defined networking (SDN) is a great solution for monitor, manage and control the network status. Opendaylight is one of the popular organizations which committed to promote SDN development. It developed a controller with the same name which named OpenDayLight (ODL). When SDN has been a core network, we are able to know anything happened in any network equipments as long as they do support Openflow protocol that is shown in Fig. 2. We can see that there is a topology in ODL GUI which includes connection, MAC and so on of linked switches and linked end devices.

The switch must be a openvswitch (OVS). Because general switch does not support Openflow protocol so that it cannot recognize the flow entry of Openflow protocol. OVS is able to easy control the switch by software way that including setting of VLAN, bridge and network interface. Figure 3 shows an example of our configuration. We set eth0 which uses out-band mode to connect controller, and then creating a bridge between eth1 and eth2. Due to SDN has not official wireless supported solution so that

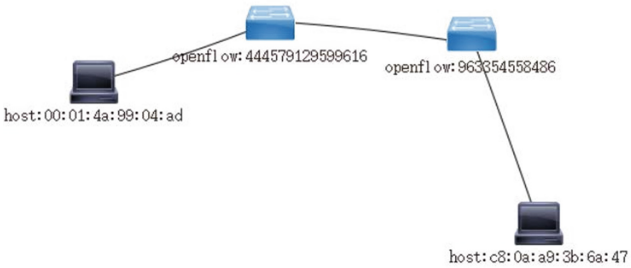


Fig. 2. Topology from ODL GUI

```
switch@switch-Lenovo:~$ sudo ovs-vsctl show
[sudo] password for switch:
4325bb34-420f-4c9e-9440-eca141e9e439
    Bridge "br0"
        Controller "tcp:192.168.5.1:6633"
            is_connected: true
        fail_mode: standalone
        Port "eth2"
            Interface "eth2"
        Port "br0"
            Interface "br0"
                type: internal
        Port "eth1"
            Interface "eth1"
```

Fig. 3. Configuration of OVS

there are not any access points (AP) can recognize packet of Openflow. In order to solve this problem, we have to implant Openwrt to AP and then mount the image file of OVS. Regrettably, current controller still cannot identify the connection which adopts to wire or wireless link. But this point is in an acceptable range.

In order to verify this platform can be interfaced with others commercial platform as well as supporting large scale environment, we purchased a HP 2920 switch as an edge switch and a pure OVS. The entire network topology is shown in Fig. 4.

After testing by wireshark, we can confirm that connectivity of this platform is no problem for data transmission than is shown in Fig. 5. Because the traffic flow in SDN must to follow the flow entry, this figure shows that a catched packet by wireshark. We can see that this packet did use Openflow protocol and successfully deliver to destination.

According to current results, we know that we already have ability to transmit packet and monitor network status through Openflow. It means that we have completed basic “perception” function. Generally, bacteria will start to move to suitable host after perception is completed. Therefore, we must to use these collected information to create the stable path for data transmission. Service Function Chain (SFC) is a project of Opendaylight. They consider that a transmission process is a collection which includes changes of service functions (SF). These service functions can be a firewall, NAT or balancer and so on possible network function. Due to each bacteria will perceive the environment when it just parasitized a host and then it will find the most advantageous way such as how to make interaction with other bacterias that including

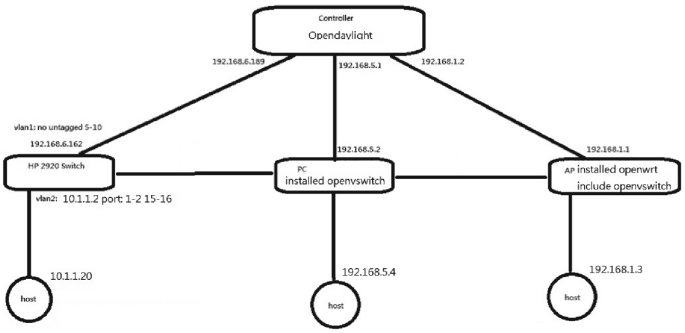


Fig. 4. Whole topology

```
Ethernet II, Src: QuantaCo_3b:6a:47 (c8:0a:a9:3b:6a:47), Dst: SonyCorp_99:04:ad (00:01:4a:99:04:ad)
  Destination: SonyCorp_99:04:ad (00:01:4a:99:04:ad)
  Source: QuantaCo_3b:6a:47 (c8:0a:a9:3b:6a:47)
  Type: IPv4 (0x0800)
Internet Protocol Version 4, Src: 192.168.5.4, Dst: 10.1.1.20
Transmission Control Protocol, Src Port: 45824 (45824), Dst Port: 6633 (6633), Seq: 1, Ack: 1, Len: 140
OpenFlow 1.3
  Version: 1.3 (0x04)
  Type: OFPT_PACKET_IN (10)
  Length: 140
  Transaction ID: 0
  Buffer ID: OFP_NO_BUFFER (0xffffffff)
  Total length: 98
  Reason: OFPR_ACTION (1)
  Table ID: 0
  Cookie: 0x2b00000000000000
  Match
  Pad: 0000
  Data
```

Fig. 5. Connectivity by wireshark

```
INFO:sfc/sfc_agent.py:Received request for SF creation: SF1
INFO:sfc/sfc_agent.py:Received request from ODL to create SF ...
INFO:sfc.common.launcher:Starting Service Function: SF1
INFO:sfc.common.launcher:Starting FIREWALL serving as SF1 at 192.168.5.2:40001,
service type:firewall
INFO:sfc.common.launcher:Starting control UDP server for SF1 at 192.168.5.2:6000
INFO:sfc/sfc_agent.py:Received request for SF creation: SF2
INFO:sfc/sfc_agent.py:Received request from ODL to create SF ...
INFO:sfc.common.launcher:Starting Service Function: SF2
INFO:sfc.common.launcher:Starting DPI serving as SF2 at 192.168.5.2:40002, servi
ce type:dpi
INFO:sfc.common.launcher:Starting control UDP server for SF2 at 192.168.5.2:6002
INFO:sfc.common.launcher:Starting Service Function Forwarder: SFF1
INFO:sfc.common.launcher:Starting SFF serving as SFF1 at 192.168.5.2:4789, servi
ce type:sff
INFO:sfc.common.launcher:Starting control UDP server for SFF1 at 192.168.5.2:600
1
INFO:__main__:Sending VXLAN-GPE/NSH/IPv4 packet to SFF: ('192.168.5.2', 4790)
INFO:__main__:Received packet from SFF: ('192.168.5.2', 4790)
```

Fig. 6. A process of SFC

competition and mutualism. It means that only creating a route is not enough. We think that the needed network functions are necessary to consider. Therefore, in the next step, we will use SFC to define the stable routing according to the needed network function or the services of any packets. These function or services are regarded as SF of SFC that are shown in Fig. 6.

4 Conclusion

In the future, network traffic will become very huge since the numbers of mobile devices are increased. Hardware solution has been unable to meet this situation completely. In this paper, we used cloud computing architecture to provide flexible adjustment and higher computing power, as well as use SDN to provide global monitoring and management. Basing on this structure which supplies a basic network function then we can implant the bacteria-inspired concept to initialize SDWBIN platform successfully. Any network traffic will able to perceive whole network status to find a most stable transmission way for packet delivery. In the future, we will try to implement "infection" and "diffusion" functions according to the definition of SDWBIN.

Acknowledgments. This research was partly funded by the National Science Council of the R. O.C. under grants MOST 104-2221-E-197-014- and 105-2221-E-197-010-MY2.

References

1. Tsai, C.-W., Cho, H.-H., Shih, T.K., Pan, J.-S.: Metaheuristics for the deployment of 5 G. IEEE Wireless Commun. **22**(6), 40–46
2. Tseng, F.-H., Chou, L.-D., Chao, H.-C., Wang, J.: Ultra-dense small cell planning using cognitive radio network toward 5 G. IEEE Wireless Commun. **22**(6), 76–83
3. Cho, H.-H., Lai, C.-F., Shih, T.K., Chao, H.-C.: Integration of SDR and SDN for 5 G. IEEE Access **2**, 1196–1204
4. Tseng, F.-H., Cho, H.-H., Chou, L.-D., Shih, T.K., Chao, H.-C.: A bacteria-inspired network platform for 5 G mobile communication system. In: The 7th International Conference on Internet (ICONI 2015), 13–16 December 2015
5. Ahmed, A.M., Xia, F., Yang, Q., Liaqat, H.B., Chen, Z., Qiu, T.: Poster: bacteria inspired mitigation of selfish users in ad-hoc social networks. In: 15th ACM International Symposium on Mobile Ad Hoc Networking and Computing (2014)

An Extension of QSL for E-voting Systems

Yuan Zhou, Hongbiao Gao, and Jingde Cheng(✉)

Department of Information and Computer Sciences,
Saitama University, Saitama 338-8570, Japan
{shuugen, gaohongbiao, cheng}@aise.ics.saitama-u.ac.jp

Abstract. E-voting is an electronic way to provide voting processes beginning from preparing ballots, following by authenticating voters and candidate registrations, through casting votes, and ending to tallying and declaring collected answers. Nowadays, there are many kinds of e-voting systems implemented to provide e-voting services over the Internet. However, there is no ad hoc method to cover the gap caused by difficult communications. QSL is a specification language for e-questionnaire systems that serves as a communication tool for specifying e-questionnaires and e-questionnaire systems. QSL is an ideal candidate because of similar processes between e-questionnaire and e-voting. The current version of QSL is reckoned without e-voting and e-voting systems. This paper proposes an extension of QSL for specifying e-voting and e-voting systems, and presents two cases using QSL for e-voting systems to show its effectiveness.

Keywords: QSL · Specification language · E-voting system · E-voting · XML · Security · Web service

1 Introduction

Elections, referenda and polls are critical processes for appropriate operation of a modern democracy [5]. Because of high efficiency and low cost of counting votes, and accessibility and convenience for disabled voters and the voters who live in remote place, many countries began to use electronic voting (e-voting) technology. E-voting is an electronic way to provide voting processes beginning from preparing ballots, following by authenticating voters and candidate registrations, through casting votes, and ending to tallying and declaring collected answers.

Over a decade, many kinds of e-voting systems are implemented to provide e-voting services over the Internet from any place and any computer or computerized equipment connecting to the Internet. Based on law, regulation, and policy, different requirements for different types of e-voting can thus derive different e-voting systems. When the government, party, and organization want to use the existing e-voting system to do an e-voting, they must specify e-voting at first to figure out the requirements. When the existing e-voting systems cannot satisfy the requirements what they want, they should order an e-voting system. It is also need to specify the e-voting system helping to clear what are necessary, and to make a clear explanation with the implementers who are not the specialists in election. Therefore, it is necessary to provide a specification language, which can help them to create precise and adequate specifications for various e-voting and e-voting systems.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_14

QSL [2, 30] is a specification language for e-questionnaire systems that serves as a general-purpose communication tool for specifying various e-questionnaires and e-questionnaire systems with a standardized, consistent, and exhaustive list of requirements. It is an ideal candidate on account of its creative concept and similar processes between e-questionnaire and e-voting, generally manifested in steps of setting up, distributing, submitting, collecting, and counting. However, the current version of QSL is reckoned without e-voting and e-voting systems.

This paper proposes an extension of QSL for e-voting systems such that users can use QSL to specify various e-voting and e-voting systems. The rest of the paper is organized as followed: Sect. 2 gives introduces in QSL. Section 3 presents an extension of QSL for e-voting systems. Section 4 shows two cases of representative e-voting systems being specified by extended QSL to confirm that QSL is hopeful to specifying e-voting systems. The related work is presented in Sect. 5. Finally, some concluding remarks are given in Sect. 6.

2 QSL: A Specification Language for E-questionnaire Systems

QSL is a specification language for specifying various e-questionnaires and e-questionnaire systems [30]. QSL serves as a communication tool with a standardized, consistent, and exhaustive requirement list to provide services to both questioner and implementer that can make questioners clearly describe the requirements for an e-questionnaire, and implementers can also clearly understand what questioners need. In addition, implementers can implement the e-questionnaire system. QSL also supports communication between questioners and an e-questionnaire system to provide only one method for questioners. Meanwhile, QSL provides services for questioners with e-questionnaire data as only one format that can be reused. QSL has an ideal state that is a tool as a complier to automatically generate e-questionnaire system for questioner [2].

QSL is based on XML [27] to provide a well-formed structure. The grammar of QSL uses XML Schema [28] to describe the structure of the documents with a precision and conciseness. QSL provides specifications of exhaustive requirements by the method of combining the primitive elements of e-questionnaires and e-questionnaire systems. QSL has been extended to specify e-testing and e-testing systems [26].

3 An Extension of QSL for E-voting Systems

3.1 E-questionnaire and E-voting

In order to extend QSL for e-voting systems, we investigated 20 e-voting systems [1, 3, 4, 6, 8, 10–12, 14–23, 25, 29] and 119 requirements [24] of e-voting systems deduced from legal input and accepted by the lawyers.

We found the similarities between e-questionnaire and e-voting from 6 aspects, which are phase, e-paper, server, software, function, and participant. Firstly, both

e-questionnaire and e-voting are not out of phases of setting up, distributing, submitting, collecting, and counting. Setting-up is to prepare software communicating with server and e-paper needed for an event. Distributing is to distribute e-paper to respondents. Submitting is to answer e-paper and send to submitting server, usually called voting phase in e-voting. Collecting is to collect the answers from respondents. Counting is to calculate the collected answers and get results, usually called tallying phase in e-voting. Secondly, e-questionnaire and e-ballot have the extremely similar contents used to express a choice preference, collectively called e-paper. E-paper consists of settings, questions, and options. Thirdly, both e-questionnaire and e-voting need server to store the collected results and server to provide registration services for respondents. Fourthly, according to phases, they have similar software to provide communication with corresponding servers. There are 2 kinds of software, one is client-side software to communicate with submitting server, another is counting software to communicate with counting server. Fifthly, both e-questionnaire and e-voting have similar functions based on phases. For instance, both e-questionnaire and e-voting need provide function of allowing access to the server if at least two different users are logged on in submitting phase. Lastly, both e-questionnaire and e-voting have similar roles of participants, which are sponsor who organizes and supports an event, questioner who designs an e-paper, analyst who processes the collected answers, monitor who monitors whether illegal or dishonest behavior occurs or not, and respondent who answers the e-paper, usually called voter in e-voting.

Moreover, we found the differences between e-questionnaire and e-voting mainly manifested in security because e-voting is mainly used in governmental elections for the universal, equal, free, and secret suffrage. More specifically, the points of a trustworthy secure e-voting system differing from e-questionnaire system are following aspects: authentication and anonymisation. Firstly, authentication is to ensure only eligible voters may cast vote only once before storing in the e-ballot box, and must ensure the casted votes are clearly separated from the identity of the voter. For instance, in order to ensure no ineligible voter cast a vote for changing election results, e-voting systems always use tokens as an authentication technique. Secondly, anonymisation is to prevent any link between the voter and his unencrypted vote. It concerns the communication channel encryption and seal method. In addition, thirdly, e-voting needs auditing phase to verify and ensure the equality of the number of voters and votes in e-ballot boxes. Auditing provides services of recording, monitoring, and verification of audit data to make authenticity and accuracy of voting results, for the security of e-voting. Fourthly, e-voting also needs certification server to provide services to validate the voters and poll workers we will explain below to prevent any possibility of affecting election results. Fifthly, e-voting needs an auditing software to communicate with submitting system. Besides, sixthly, e-voting needs a list of candidates, and provides a candidate nomination and candidate registration. At last, e-voting also needs voting worker to help start submitting phase, make a selection on e-paper, resume submitting phase after any kind of exception, malfunction, or breakdowns, check the system state, close submitting phase, and start counting phase, especially in parliamentary elections.

3.2 Extending QSL for E-voting Systems

According to the similarities and differences between e-questionnaire and e-voting, we propose an extension of QSL for e-voting systems based on the current QSL for e-questionnaire and e-testing systems, and we change properly the structure of it. Figure 1 illustrates QSL structure of relationship overview among e-questionnaire, e-testing, and e-voting.

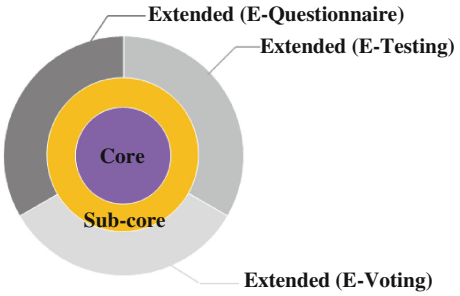


Fig. 1. Relationship overview among e-questionnaire, e-testing, and e-voting.

In order to extend and upgrade QSL easily, we design that QSL structure has 3 layers. In the innermost layer, QSL defines core elements. Specifying any e-questionnaire, e-testing, and e-voting system must specify all the core elements. The core element consists of the combinations of the elements in the middle layer. The elements in middle layer are called sub-core elements. In the outermost layer, there are 3 isolated ranges, which are for e-questionnaire, e-testing, and e-voting, respectively. The elements in this layer are called extended elements. For example, if a user wants to specify an e-voting system, he/she shall specify all the elements in the innermost layer and middle layer, and specify all or part of the elements in the outermost layer depending on the security level.

In order to well define the combinations of core elements, sub-core elements, and extended elements, we use double-digit to mark the elements. In the ten's place, 0, 1, and 2 stands for core element, sub-core element, and extended element, respectively. Table 1 shows a list of the elements with the double-digit numbers. In the one's place, 0 stands for a special mark. The elements are associated with the namespace defined using **QSL**. As the configuration of core elements, it gives a combination relationship of numbers of sub-core elements and extended elements defined only for e-voting. As well, some major elements for constructing sub-core elements are shown below.

Based on the current QSL, we exact **Security** as a core element from **System**. To well deal with the intruder's technical capabilities, from core, through sub-core, to extended elements, QSL exhibits its extensibility to easily revise elements in middle and outermost layers. We design QSL providing the phase security isolation. In other words, there are different security requirements in each phase.

In security aspect, we exact **Authentication** and **Anonymisation** as two important elements in the outermost layer. **Authentication** has 3 child elements, which are

Table 1. Core elements, sub-core elements, and extended elements for e-voting.

Group	No.	Element	Configuration
Core	00	QSL	–
	01	Security	Phase, Server, Software, Participant
	02	System	Phase, Function, Server, Software
	03	EPaper	Phase, Paper, Participant
	04	Data	Phase, Paper, Participant
Sub-core	11	Phase	SettingUp, Distributing, Collecting, Submitting, etc.
	12	Paper	Arrangement, Text, Media, Question, Option, etc.
	13	Function	Func-Import, Func-Export, Func-Distribute, etc.
	14	Server	RAServer, PSServer
	15	Software	CSSoftware, CTSofteware
	16	Participant	Sponsor, Questioner, Analyst, Monitor, Respondent
Extended (E-V)	21	Authentication	Secret, Token, Biometrics
	22	Anonymisation	Seal, Channel
	23	Auditing	–
	24	CAServer	ID, Link
	25	ATSoftware	Version, ID, Link, Solution
	26	Candidate	ID, Name, Affiliation, Proposer
	27	Admin	–

Secret, **Token**, and **Biometrics** in a choice relationship. Each child element is designed as string type and derived by restriction with enumeration values. An example is shown below. To maximize the security for difficult to fake the card, the most popular method is the combination of knowledge of a secret and ownership of a token. *TAN* is a unique code of letters and digits send by a secure post to eligible voters in election setup phase. *ID Card* is pre-existing election authentication card.

```
<Authentication>
   <Secret>TAN</Secret>
   <Token>ID Card</Token>
</Authentication>
```

As to **Anonymisation**, it has 2 child elements, which are **Seal** and **Channel**. **Seal** has an attribute to specify the seal method and a content to specify what is sealed. To ensure secure communication channel, the value of the element contains “Internet Link”, “Phone”, “E-mail”, and so on. **Channel** has an optional attribute named *ID*. The optional attribute *Method* belongs to **Anonymisation** to specify the method containing “BlindSignature”, “SeparationOfDuty”, “HardwareSecurityModel”, etc.

```
<Anonymisation>
   <Seal Method="Signature">Vote</Seal>
   <Channel>Internet Link</Channel>
</Anonymisation>
```

In order to specify different servers in an event for their different functions and authority, in essence for security, **Server** has 2 groups, one is called “ServerGroup”,

another is called “SeparateServerGroup”. The first group is used in a situation that all the phases are executed on a server, which has 2 child elements named **ID** and **Link**. It is used to deal with the communication links between software and servers. Alternatively, the second group is used to separate into 3 servers according to the phases. In this group, there are 3 child elements, which are **RServer** (Registration Server) to let voter register, **CAServer** (Certification Server) to confirm whether the voter is eligible or not, and **PSServer** (Paper Storage Server) to store the e-votes in e-ballot boxes. Each child element has a child group named “ServerGroup” in order to connect with corresponding servers. For example, in below situation, the registration server named “Re1” links to a software named “Soft1”.

```
<RServer>
   <ID>Re1</ID>
   <Link Ref="SoftID">Soft1</Link>
</RServer>
```

Corresponding to server and system, **Software** has 2 groups, one is called “SofrwareGroup”, another is called “SeparateSoftwareGroup”. The first group is used in a situation that all the phases are executed using software, which has an optional child element named **Version**, and 3 necessary child elements named **ID**, **Link**, and **Solution** in sequence. The second group is used to separate into 3 kinds of software according to the phases for security, especially in auditing phase. In this group, there are 3 child elements, which are **CSSoftware** (Client-side Software) to communicate with submitting system, **CTSoftware** (Count Software) to communicate with counting system, and **ATSoftware** (Auditing Software) to communicate with auditing system. Each child element has a child group named “SoftwareGroup” in order to connect with corresponding systems. For example, the software named “Soft1” links to an e-voting system named “Sys1”, and uses web browser to perform an e-voting.

```
<Software>
   <ID>Soft1</ID>
   <Link Ref="SysID">Sys1</Link>
   <Solution Type="Web Browser"/>
</Software>
```

We add **Auditing** as the last phase, in the sequence of child elements of **Phase**. It can combine with **Server** and **Software**. In addition, we add **Admin** as a role of poll worker, in the sequence of child elements of **Participant**. Because poll worker has different duties in an event to prevent bribe, **Admin** has an attribute *Role* to distinguish each duty related to authority. Besides, we add **Candidate** and its child element, which are **ID** to identifier, **Name** as necessary information, **Affiliation** as optional information, and **Proposer** related with nomination. The candidate information is used as an option of the question. The nomination is similar to a questionnaire proposed by a group of respectable people, is used to get candidate list as question options. It can combine with **RAServer** to specify candidate registration.

4 The Cases of Using QSL for E-voting Systems

We used extended QSL to specify the Estonian system, which was used to hold a federal election in Estonia in 2007 [7]. In addition, we also specify POLYAS system [13], a famous voting system founded in 1996 has been used to cast more than one million vote a year for its elections. As the results, the extend QSL can specifying the whole requirements, especially all the security requirements. In other words, the elements in QSL structure for e-voting systems can be used to specify these 2 systems. As examples, two representative security specification snippets for submitting phase of Estonian System and setting up phase of POLYAS System by QSL.

Estonian System uses a combination of possession-based and secret-based authentication method. The voter used the first secret key, an ID card, to identify the authority of casting a vote and the second secret key to sign his encrypted vote. Voter cannot change his vote after casting a vote. In submitting phase, the voter chooses and the system shows his choice to let voter to verify or change his choice again before submitting vote to the e-ballot box. When he submits his confirmed vote, voter uses his second secret key "Signature" and a blind signature scheme to encrypt the vote. In this snippet, it specifies the authentication method named "Signature" the voter has, a secure Internet link, and a sealed and encrypted anonymous method. In addition, only an eligible voter can cast a vote. If any kind of exceptions, malfunction, or breakdown occurs, the certified poll workers can resume this phase. The certified monitor can monitor the phase. Both worker and monitor cannot authority to see the votes.

```
<Submitting>
   <Authentication>
       <Biometrics>Signature</Biometrics>
   </Authentication>
   <Anonymisation>
        <Seal Method="Signature">Vote</Seal>
        <Channel>Internet Link</Channel>
   </Anonymisation>
   <Authority>
      <Participant>
        <Monitor Role="Certified" Situation="Phase"/>
        <Admin Role="Certified" Situation="Phase"/>
        <Respondent Role="Certified"
Situation="Ballot"/>
      </Participant>
   </Authority>
</Submitting>
```

POLYAS System uses the secret-based authentication method named "TAN", and the separation of duty approach to anonymise voter's identity among multiple servers. After logging on client-side voting software, voter uses his secret key through the first directed SSL connection from the software and registration server. The server checks whether the requesting voter is eligible or not. In order to ensure one voter one vote, the ID code is sent to the vote storage server through the second directed SSL connection,

which checks whether the voter has already cast a voter. If this voter has not vote, the registration server generate a ballot belongs to this particular voter.

```
<SettingUp>
  <Authentication>
    <Secret>TAN</Secret>
  </Authentication>
  <Anonymisation Method="SeparationOfDuty">
     <Channel ID="SSL1">Internet Link</Channel>
     <Channel ID="SSL2">Internet Link</Channel>
  </Anonymisation>
  <Authority>
    <Participant>
      <Respondent Role="Certified"/>
    </Participant>
  </Authority>
</SettingUp>
```

5 Related Work

EML is a standard for the structured interchange of data among hardware, software, and service providers who engage in any aspect of providing election or voter services to public or private organizations [9]. EML is a uniform and reliable method to allow systems supporting the election process to interoperate. EML is based on XML.

In essence, QSL has different purpose, method, and scope in contrast with EML. Firstly, QSL is aimed at as a communication tool among the questioner, implementer, and the system to fill the gap caused by the difficult communications. QSL serves as a formalized specification to QSL compiler automatically generate e-questionnaire, e-testing, and e-voting systems. As the purpose, EML is focused on defining open, secure, standardized and interoperable interfaces between components of election systems for data exchange. EML gives a detailed structure how to transmit information helping to implement the e-voting system. Secondly, QSL is designed through the combinations of a small amount of primitive elements for more comprehensive requirement specifications. EML is designed with a terminology based on two complementary high-level process models of an election exercise. Thirdly, to distinguish from scope, QSL is used to specify e-voting systems, as well as e-questionnaire and e-testing systems. EML can only specify e-voting systems.

6 Conclusion

In this paper, we have proposed an extension of QSL for various e-voting and e-voting systems, presented two security specification snippets of Estonian System and POLYAS system by QSL. From the specifications of these 2 systems, we confirm that QSL is hopeful to specify e-voting systems.

The extended QSL can provide convenience for questioner to design e-voting systems, to communicate clearly with implementer with several desirable necessary specifications, and to reuse the data with a unified format; and also provide convenience for implementer to clearly understand what questioner needs even the implementer are not the specialist in election. Moreover, the extended QSL emphasizes the specifications of security requirements, which can be used to further improve the specifications for e-questionnaire and e-testing systems.

In the future, we will continue working on improving QSL for various e-voting and e-voting systems, and implementing a compiler of QSL to provide convenience to automatically generate e-questionnaire, e-testing, and e-voting systems.

References

1. AddPoll. http://www.addpoll.com
2. AISE Lab, Saitama University: QSL Manual (ver. 1.5) (2014). http://www.aise.ics.saitama-u.ac.jp/QSL
3. FC2 Vote (in Japanese). http://vote.fc2.com
4. Google Form. http://docs.google.com
5. Grizalis, D.A. (ed.): Secure Electronic Voting. Advances in Information Security, vol. 7. Kluwer (2002)
6. Helio. https://vote.heliosvoting.org
7. Madise, U., Martens, T.: E-voting in estonia 2005. The first Practice of country-wide binding internet voting in the world. In: Krimmer, R. (ed.) Proceedings of the 2nd International Workshop on Electronic Voting 2006, LNI GI Series, Bonn, Germany, pp. 15–26 (2006)
8. Mobo Survey. http://www.mobosurvey.com
9. OASIS Election Markup language (EML) Specification Version 7.0 (2010). http://docs.oasis-open.org/election/eml/v7.0/cs01/eml-v7.0-cs01.html
10. OQSS. http://www.oqss.com
11. Opinion Stage. http://www.opinionstage.com
12. Opoll (in Chinese). http://www.opoll.com
13. POLYAS. https://www.polyas.com
14. Poll Every Where. http://www.polleverywhere.com
15. ProProfs. http://www.proprofs.com
16. Qadah, G.Z., Taha, R.: Electronic voting systems: requirements, design, and implementation. Comput. Stand. Interf. **29**(3), 376–386 (2007)
17. QuestionPro. http://www.questionpro.com
18. Simple Voting. https://www.simplyvoting.com
19. Snappy Poll. https://www.snappypoll.com
20. Stone Poll (in Chinese). http://www.stonepoll.com
21. SurveyMonkey (in Japanese). http://www.surveymonkey.com
22. Tou Piao Wang (in Chinese). http://www.toutoupiao.com
23. Voice Poll. https://voicepolls.com
24. Volkamer, M.: Evaluation of Electronic Voting. LNBIP, vol. 30. Springer, Heidelberg (2009). doi:10.1007/978-3-642-01662-2
25. Votenet. http://www.votenet.com

26. Wang, Z., Zhou, Y., Wang, B., Goto, Y., Cheng, J.: An extension of QSL for e-testing and its application in an offline e-testing environment. In: Park, J.J., Chao, H.-C., Arabnia, H., Yen, N.Y. (eds.) Advanced Multimedia and Ubiquitous Engineering. LNEE, vol. 352, pp. 7–14. Springer, Heidelberg (2015)
27. W3C: Extensible Markup Language (XML) 1.0, 5th edn. http://www.w3.org/TR/2008/REC-xml-20081126/
28. W3C: XML Schema. https://www.w3.org/XML/Schema
29. YouPoll. http://www.youpolls.com
30. Zhou, Y., Goto, Y., Cheng, J.: QSL: A Specification Language for E-questionnaire Systems. In: Proceedings of the 5th IEEE International Conference on Software Engineering and Service Science (ICSESS 2014), Beijing, China, pp. 224–230. IEEE (2014)

Behavior-Based Detection for Malicious Script-Based Attack

Soojin Yoon(✉), Hyun-lock Choo, Hanchul Bae, and Hwankuk Kim

Information Security R&D Technology Sharing Center,
KISA (Korea Internet & Security Agency), 17F, IT Venture Tower,
135 Jungdae-ro, Songpa-gu, Seoul, Korea
{sjyoon, hlchu, hcbae, rinyfeel}@kisa.or.kr

Abstract. Several DoS attacks have occurred through web browsers, not from malicious executable files. Most tools used in web attacks are downloaded malware. As the dynamic functions of HTML5 can be performed on a web browser, however, the latter can be abused as an attack tool. The features of web browser-based attacks are different from those of previous attacks, so a different detection method is needed for malicious behavior on web browsers. This paper introduces script-based attacks made through web browsers, and proposes a detection method based on a web browser's behavior.

Keywords: Web · Web attack · Web browser · Script-based attack · HTML5

1 Introduction

SohuTV (2014) [1], GibHub (2015) [2] and IMGUR (2015) [3] were hit by massive traffic from DoS (denial of service) attacks. The common thing among these attacks was that they were script-based attack.

A script-based attack is composed of pure Javascript and HTML elements and runs on a web browser. Most web attacks are drive-by-download, which is inducing users to download malware without thought. A script-based attack is different, and this paper explains the details of script-based attacks in Sect. 2.

Because of the distinct features of script-based attacks, typical detection methods do not work. This paper proposes a new detection method for script-based attacks on web browsers.

2 Relative Works

2.1 Script-Based Attack

A script-based attack acts as malicious behavior on a web browser. For example, SohuTV, GibHub and IMGUR were attacked by massive traffic from users' browsers. It used functions to send traffic to users' web browsers. The functions used were innocent and did not constitute malicious behavior. But when attackers replayed it, users did not know of their attacks.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_15

The big difference is the attack tool. Drive-by-download's tool is an executable file. So prevention lies with detecting a downloaded file and checking that the file is malicious. But a script-based attack runs a webpage through a user's web browser with no file downloaded. Moreover, after closing the web browser, no evidence remains in a user's device except that they connected. Detecting web pages including malicious script is not unhelpful. In addition, as HTML5 appeared and added new functions on the web, the influence of script-based attack will grow stronger [4].

2.2 Previous Detection Method

Among the distinct features of a script-based attack, one detection method stands out.

Spyproxy [5] and WebShield [6] run on proxy servers between web servers and users to protect users. Spyproxy catches a web page that a user connects to and runs it though a VM worker. The VM worker detects behavior such as making a new process, manipulating a file or breaking a sandbox. Spyproxy also picks users' action on a web page before sending it to a web server. Spyproxy tests selected user actions on VM worker and whether it is safe. Spyproxy can detect malicious web pages, but the criteria of detection focuses on drive-by-download attacks. WebShield also uses sandbox to run a web page. The major difference from Spyproxy is that a web browser in sandbox makes a DOM structure and shares it with a user's browser instead the browser making the DOM structure from a web server's response.

JSand [7] is specialized for third-party Javascript. It provides a platform to run third-party Javascript through JSand sandbox environment. It blocks the outer actions of third-party Javascript that can cause malicious behavior or side effects, and is for web developers.

Our team has developed a method to detect script-based attacks. Signature-based [8], network-based (proxy server) [9] and web site checker [10] are proposed. This paper expands on the findings of a previous paper [11].

3 Overall of Detection Method

The proposed method has three steps. First is running a webpage through a virtual browser. A virtual browser is based on a web browser and runs the functions of the webpage and reports the running functions. In this paper, the reported functions are called behavior nodes, which indicate functions including parameter and related information. In Capture 4, the behavior node is explained in detail. A virtual browser also blocks massive traffic from the running of a webpage because it could be a DoS attack. Second is the building behavior node tree. After writing behavior nodes, the final step is distinguishing between malicious and innocent using FSM criteria.

4 Virtual Browser

We implemented a virtual browser using Webkit. The first of its two parts consists of that for running a webpage and blocking massive traffic that can cause leaks of user information or a DoS attack. The second is for controlling a previous webpage and recording its behavior.

The control part lets the running part play the events of a webpage. Events of a webpage are actions like clicking images, inserting values into input boxes and so on. We gathered events, selected the most used events and made a list of them. Our virtual browser ran events on the list to save time and resources.

The control part records the behavior of a webpage as a result of running. Behavior is classified into three categories: JS API, DOM and EVENT.

JS API is behavior by Javascript API, as it is written with API, parameter value and object of API. For example, sending of an XMLHttpRequest, creator worker of Worker can be classified as JS API. DOM is an element of a webpage like an image or iframe. EVENT is triggered by a user or dynamic function. Onclick is directed to users when they click on something and onfocus is active when an element is clicked or selected. They are EVENT.

The order of recording follows that of running or playing.

4.1 Behavior Node

Behavior node includes the behavior of a webpage and relative information about it. Our system transmits information on behavior into a behavior node as set. The written information is Table 1.

Table 1. Information of behavior node

Name	Explanation
NodeID	Unique number for behavior node, increases when adding new nodes
PageID	Unique number for webpage including behavior node
type	Type of behavior node, one of EVENT, JS API and DOM
Detail	EVENT, JS API or DOM element of node
	Ex) window.open()→open, <img onclick = myfun()>→(DOM Node)img, (EVENT Node)onclick
Object	(EVENT Node) tag information related to event node
	(JS API Node) object information related to JS API node. As window object is omitted in webpage, JS API nodes of window object can omit object information
Ancestor	(Be Added when making tree) causes behavior node
Depth	(Be Added when making tree) Depth of behavior node in behavior tree
Value	Assigned value of event node. Ex) <img onclick=myfun()..>→myfun
Date	The date of recording the behavior node. YYYYMMDDHHmmssffffff

When writing a behavior node in a virtual browser, ancestors and depth are exceptions. They can be added through building behavior tree. One behavior node might have several parameters and attributes. Our system records parameters and attributes separately.

5 Behavior Node Tree

The behavior node tree is a graph that connects behavior node and shows flow among behaviors. The behavior node tree builds in the order of times that behavior occurred (Fig 1).

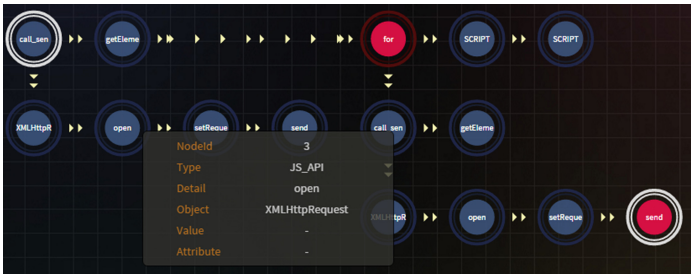


Fig. 1. Behavior node tree

There are two methods for connecting behavior nodes. One is horizontal connection, which just connects a behavior node to a previously occurring behavior and post-occurring behavior. The other is vertical connection. The upper node is the cause and lower node the result. As there is a vertical connection, nodes of the behavior node tree each have their own depth.

Horizontal connection is default. When the JS API node follows the EVENT node, they have a vertical connection. The EVENT node causes the JS API node as in the following. Our system builds a behavior node tree through these connections.

6 Malicious Detection

The behavior node tree is an input of malicious detection. The criteria of detection are element, order and density of a behavior node. It is formed as an FSM (Finite State Machine) and written in XML.

The transition condition of states is element or depth condition of a behavior node. Pure FSMs have no recording, but we have depth recording in this system.

An FSM traces the behavior node tree from vertical connection to horizontal connection.

The Fig. 2 shows criteria of CSWSH. First, state checks where behavior node is JS API that have WebSocket objects and other domains as parameters. If the condition is met, record the depth of the node and the state is transferred to state 2 from 1. State 2

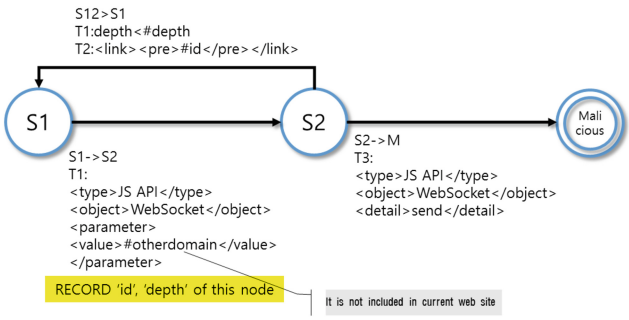


Fig. 2. FSM for CSWSH (Cross Site WebSocket Hijacking)

has two conditions to transfer. The backward condition is that depth of the new node is lower than or the same as the recorded one. The forward condition is checking the JS API node that has sent API. The final state is malicious and end state. If an FSM reaches the malicious state in tracing the behavior node tree, the webpage writing of that tree will perform malicious action.

An FSM checks malicious elements, running order and relations between cause and effect, and is helpful in detection of malicious webpage delicately.

We wrote 23 FSMs based on attacks using HTML5 (Table 2).

Table 2. Detection to various HTML5 attacks

	Malicious type	Original	Rename	Code split
1–9	Hash DoS/Port Scan/XML DoS (setInterval)/ CSWSH/Drag&Drop/WebSocket Information Leakage/Cross Site Printing/Geolocation Leakage/IndexedDB Leak/Script DoS (for loop)	O	O	O
10–16	Click info/WebStorage Leak/History Modification/Vibration Attack/Client DoS/AutoComplete Leakage	O	O	No Sample
17	Cookie Sniffing	X	X	X
18	Network Scan	Error	Error	X
19	SVG Keylogger	X	X	No Sample
20–21	MouseLogger/Server-Sent Event Bot	O	O	X
22–23	Worker DDoS/Recursion by SVG	Error	Error	Error
Detection Rate		90 %	90 %	77 %

7 Experiment

7.1 Detection Rate

We tested 23 types of attacks with three sample types. The first sample was original with no manipulation. The second was a rename code that changes the name of variables and functions from the original code. The last one was the split code used to split APIs in two or more files. No sample means making the type of sample is hard.

Our system detected malicious samples in 90 percent except errors. We guess the cause of errors is sending massive traffic or using SVG. Our virtual browser blocked massive traffic to stop attacks in advance so it might stop writing behavior nodes. Parsing SVG needs an additional SVG engine, but our browser had none to parse other media. SVG is a vector image and written XML.

7.2 False Alarm

We gathered safe websites to test for false alarms. We picked 50 sites from the ALEXA top 100 [15] that passed Google Safe Browsing [16] and Norton Safe Web [17]. One of them was detected as script DoS. The reason was that it sent its web log to other domains periodically. But because the period was long, DoS seldom occurred. We should add elements to lessen false alarms.

8 Conclusion

The proposed method is detecting malicious behavior in on a webpage through running it. A virtual browser provides a safe environment for analysis. Behavior from a webpage is defined as a node and arranged to build a tree structure. An FSM detects malicious behavior expressed as a behavior node tree, which is more delicate than using pattern matching.

Our virtual browser can run webpages and record behavior but is limited in that it is hard to run SVG or additional web elements. Moreover, the frequency or effects from behavior are not criteria to detect. We will improve on these limits and add solutions for other web-based attacks to make our system a total web solution.

Acknowledgments. This work was supported by the ICT R&D Program of MSIP/IITP. [B0101-15-0230, The Development of Script-based Cyber Attack Protection Technology]

References

1. One of World's Largest Websites Hacked: Turns Visitors into 'DDoS Zombies', INCAPSULA. https://www.incapsula.com/blog/world-largest-site-xss-ddos-zombies.html
2. GitHub jammed by injected JavaScript, servers whacked by DDoS, The Register. http://www.theregister.co.uk/2015/03/27/github_under_fire_from_weaponized_great_firewall/

3. Imgur suffers DDoS attack on 4chan and 8chan servers, SCmagazine. http://www.scmagazine.com/imgur-suffers-ddos-attack-on-4chan-and-8chan-servers/article/440522/
4. Yoon, S., Jung, J.H., Kim, H.K.: Attacks on web browsers with HTML5. In: ICITST 2015 (2015)
5. Alexander, M., et al.: SpyProxy: Execution-based Detection of Malicious Web Content. USENIX Security (2007)
6. Zhichun, L., et al.: WebShield: enabling various web defense techniques without client side modifications. In: NDSS (2011)
7. Pieter, A., et al.: JSand: complete client-side sandboxing of third-party JavaScript without browser modifications. In: Proceedings of the 28th Annual Computer Security Applications Conference. ACM (2012)
8. Yoon, S., et al.: Automatic attack signature generation technology for malicious javascript. In: ICMIC (2014)
9. Choo, H., et al.: The analysis engine for detecting the malicious javascript. In: ICOINI (2014)
10. Bae, H., et al.: Study on inspection of website vulnerability and risk assessment method. In: World IT Congress 2016 (2016)
11. Choo, H., et al.: The behavior-based analysis techniques for HTML5 malicious features. In: IMIS 2015 (2015)
12. https://www.w3.org/TR/html51/
13. http://www.w3schools.com/tags/ref_eventattributes.asp
14. https://developer.mozilla.org/en-US/docs/Web/Guide/HTML/HTML5
15. http://www.alexa.com/topsites/global;0
16. https://developers.google.com/safe-browsing/
17. https://safeweb.norton.com/

The SP-tree: A Clustered Index Structure for Efficient Sequential Access

Guang-Ho Cha(✉)

Department of Computer Engineering, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea
ghcha@snut.ac.kr

Abstract. We introduce the *SP-tree* that is a variant of a multidimensional index structure, with the object of offering efficient sequential disk access. The SP-tree is based on the index clustering technique called the *segment-page clustering* (*SP-clustering*). Most relevant index pages are widely scattered on a disk due to dynamic page allocation, and thus many random disk accesses are required during the query processing. The SP-clustering avoids the scattering by storing the relevant nodes contiguously in a *segment* that contains a sequence of contiguous disk pages and improves the query performance by offering sequential disk access within a segment. Experimental results demonstrate that the SP-clustering improves the query performance up to several times compared with the traditional ones with respect to the total elapsed time.

Keywords: Index clustering · Sequential disk access · Multidimensional index

1 Introduction

More than dozens of years of database research have resulted in a great variety of multidimensional indexing methods (MIMs). However, traditional MIMs tend to randomly access many index pages because the index pages are widely scattered on a disk due to dynamic page allocation. To avoid the performance degradation due to many random disk accesses, the related index nodes need to be clustered. However, existing MIMs do not take into account the clustering of *indices*. They take into consideration only the clustering of *data*. Moreover, the *dynamic* index clustering requires the on-line index reorganizations, and the overhead of the global index reorganization is excessive.

To overcome the drawbacks of the existing multidimensional indexing methods, we propose the *segment-page clustering* (*SP-clustering*) technique. The SP-clustering is based on the concept of *segments*. It considers the disk to be partitioned into a collection of segments. Each segment consists of a set of L contiguous pages on disk. A segment is the unit of clustering in the SP-clustering. All disk pages in a segment can be read by a single disk sweep, and thus it saves much disk startup and seek time. In the SP-clustering, all disk pages are addressed by a pair of (segment no, page no).

This research was supported by Basic Science Research Program through the National Research Foundation of Kore (NRF) funded by the Ministry of Education (NRF-2014R1A1A2059306).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_16

This addressing scheme allows that pages as well as segments can be used as the disk access units. When random accesses are required or query ranges are very small, page-based disk accesses can be used instead of segment-based accesses.

2 Related Work

In the literature, some techniques for sequential disk access have been proposed. However, most of the work focuses on how to cluster data on a disk or how to physically maintain the data in sequence, and less attention has been paid to the clustering or sequential accessing for indexes.

The concept of the segment is similar to the idea of the *multi-page block* used in the SB-tree [5] and the bounded disorder (BD) access method [3, 4], which are variants of the B-trees, in the sense that they accommodate a set of contiguous pages and support multi-page disk accesses. However, this concept has not been applied to MIMs because it might consume the disk bandwidth excessively with increasing dimensionality. As an instance, let us suppose that a query range overlaps only a half on each dimension of the data region occupied by a segment. Then the wasteness of the disk bandwidth caused by reading a segment instead of reading individual pages is ½ (= 1 − ½) in one-dimensional case, while it is $1 - (½)^d$ in d-dimensional case. In fact, however, the multi-page disk reads such as segment reads are more needed in high dimensions because the probability that the query range overlaps with the regions covered by the index nodes increases with the dimensionality due to the sparsity of the domain space, and thus more disk pages are required to be read in higher dimensions. In addition, unlike the multi-page blocks used in the B-trees in which all index nodes as well as all data objects have total ordering among themselves, the index nodes within segments for MIMs have no linear order among them. This makes the design and maintenance, such as partitioning and merging, of the segments in MIMs more difficult than those of the multi-page block in the B-trees.

3 The SP-clustering

In this section we introduce the SP-clustering. To demonstrate the effectiveness of the SP-clustering, we apply it to the LSD-tree [1, 2] and call the resultant index tree the *SP-tree*. We should note that our focus is the SP-clustering not the SP-tree although we explain the SP-clustering through the SP-tree. The SP-clustering technique can be applied to most MIMs including the R-trees.

3.1 The Structure of the SP-tree

The SP-tree is a multidimensional index structure to index d-dimensional point data, and its underlying structure is the LSD-tree. We chose the LSD-tree for implementation of the SP-clustering because we were dealing with d-dimensional point data and the LSD-tree has high fanout. The SP-tree considers the disk to be partitioned as a

collection of segments. Segments are classified into *nonleaf segments* and *leaf segments*. The nonleaf segment accommodates nonleaf nodes of the index tree and the leaf segment holds the leaf nodes. The reasons why we separate the segments into two kinds are twofold: it simplifies the design of the index structure and it encourages the upper part of the index structure to reside in the main memory when we cache the index into the main memory. We call the nonleaf segment *n-segment* and the leaf segment *l-segment*.

Each segment consists of a set of f (e.g., =32) contiguous pages on disk, that can be read or written with a single sweep of the disk arm. From the first page encountered by the disk head in reading and writing a segment to the last one in the segment, the pages are numbered 1, 2, …, f. A segment has the following properties:

- A segment consists of a set of f nodes which reside on contiguous pages on disk. The number f is called the *fanout* of a segment.
- k, $1 \leq k \leq f$, nodes falling in a segment are filled contiguously from the beginning of the segment.
- The SP-tree reads k nodes from a segment at a time rather than all f nodes, and thus it saves the disk bandwidth.
- Every node of the SP-tree sits on a segment.
- Leaf nodes reside in an l-segment and nonleaf nodes are in an n-segment.

Consider Fig. 1. A node at level j-1 (level 0 is the root) contains pointers (ptr 1, ptr 2,…, ptr k) to child nodes (pages) and the separators ((dim 1, pos 1), …, (dim k, pos k)). A separator contains a pair of a *split dimension* and a *split position* in the dimension. The entry M denotes the number of entries in the node. The contiguous sequence of page numbers, ptr 1 through ptr k, in the node at level $j - 1$ points to nodes at level j, node 1 through node k, which sit on the first k pages of a single segment, i.e., $M = k$. A specific node can be accessed directly by a pair of (segment number, page number), and k nodes can be accessed sequentially by the segment number and the value of M.

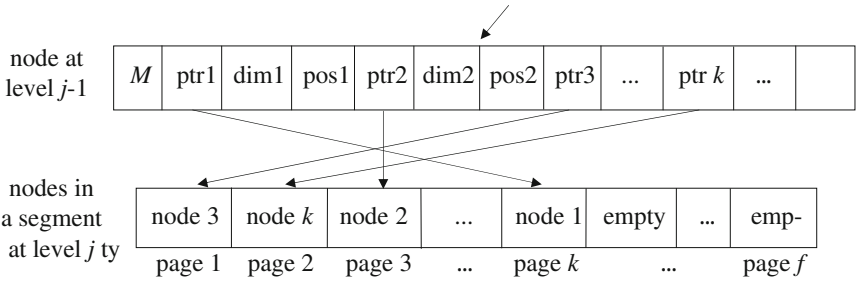


Fig. 1. Contiguous pages on a segment

3.2 Building the SP-Tree

The SP-tree has a hierarchical index structure. As usual for index structures which support spatial accesses for point data, the SP-tree divides the data space into pairwise

disjoint cells. With every cell a data page is associated, which stores all objects contained in the cell. In this context, we call a cell a *directory region*.

Successive parts of Fig. 2 show how the SP-tree grows and how its nodes are clustered in a segment. When the first entry is inserted, a single page of an *l*-segment is allocated for the first node of the SP-tree. This node is a root node as well as a leaf node. The figures in the left side of Fig. 2 show the procedure of partitioning the 2-dimensional data space. We assumed that the range of each dimension is 0 to 100, and a pair of numbers on the directory regions indicates (*l*-segment number, page number).

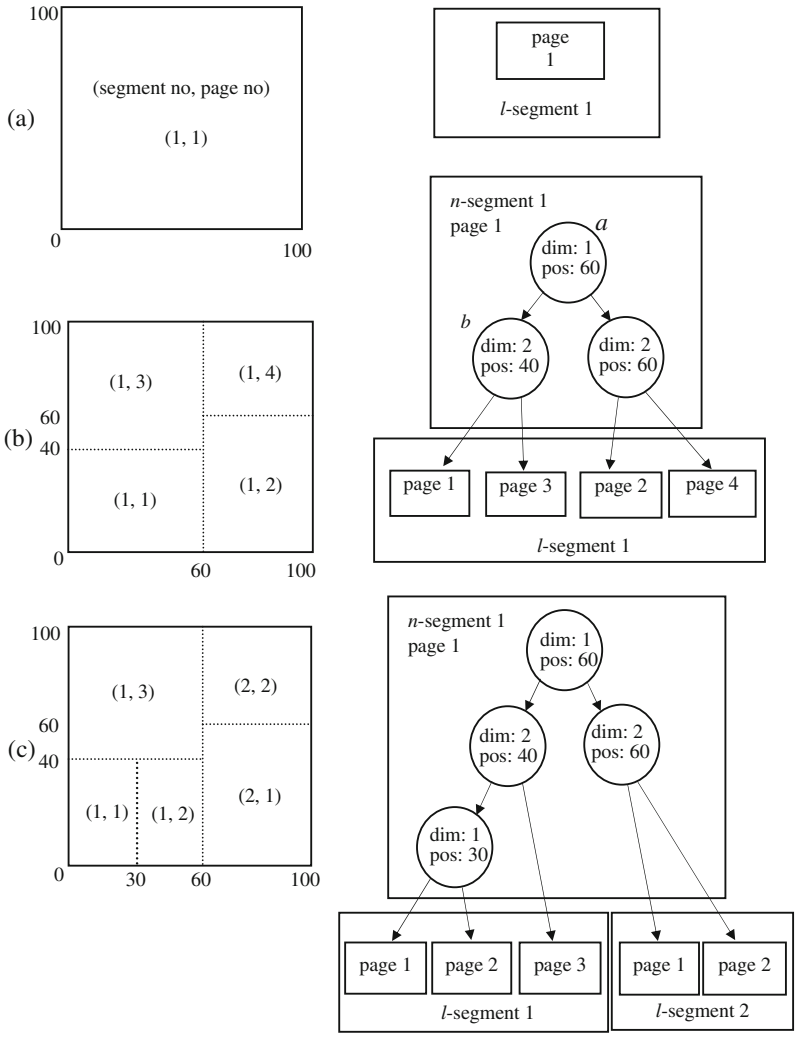


Fig. 2. The growth of the SP-tree

Successive entries are added to the node until an insert forces a split in the node. This node is then split into two leaf nodes which occupy page 1 and page 2 of the *l*-segment 1. An *n*-segment is allocated and the first page of the *n*-segment is assigned for new root node. The root node now contains a single separator and two pointers. A separator contains the information about the *split dimension* and the *split position* in the dimension. In the example of Fig. 2, the split is performed at position 60 in dimension 1. With subsequent insertions, overflows occur in the *l*-segment 1 and they cause the node split. Whenever a node split occurs, the SP-tree looks for an empty page on the segment containing the node receiving the insert. As shown in Fig. 1, this will be the page number $k + 1$ in the containing segment, where k nodes already exist. The SP-tree keeps the information in the node which tells us how many pages are occupied in each segment, i.e., M in Fig. 1. If an empty page exists, we place the new node created by the split on that page. If there is no empty page in the segment, then a *segment split* is necessary. A new segment S is allocated, and the overfull segment containing the splitting node is read into the memory. Then the $f + 1$ nodes of the segment are distributed into two segments.

3.3 Segment Split Strategy

An important part of the insertion algorithm of the SP-tree is the segment split strategy which determines the split dimension and the split position. First, the SP-tree finds the internal node u which (directly or indirectly) plays a role of root for the overfull segment R. The separator of the internal node u has the dimension and the position to split the segment R. In Fig. 2(b), for example, if a new entry is inserted into the page (1, 1) and it causes the *l*-segment 1 to overflow, the SP-tree finds the internal node which plays a role of root of the *l*-segment 1, it is the internal node a in the case of Fig. 2(b). Since the SP-tree maintains an array to save the traversal path from the root to the target page where a new entry to be inserted, it is not difficult to find the internal node that plays a role of root for the overfull segment. Starting from the root of the SP-tree, we check if the overfull segment can be split into two when we apply the separator of the current internal node to split the segment. If the segment can be split using the separator, the corresponding internal node is selected as the root node of the overfull segment, and the segment is split. The data pages belonging to the right children of u are reallocated to the front positions of a new segment S, and the remaining pages are moved forward so that they fall on the front pages of the segment R. As a result of this segment split strategy, the data pages under the same internal node are collected in the same segment.

Consider Fig. 3. It shows the effect of a nonleaf page split. In Fig. 3(a) we are to insert a new index entry into the nonleaf page P due to a page split in lower level segment S'. If the page P overflows, the SP-tree allocates a new page P' and partitions the entries in P all but the local root entry u including a new index entry into two pages P and P'. After the partition, the local root entry u that was in P is promoted to the parent page, and the two pages P and P' include the entries belonged to the left and the right subtrees of u, respectively (see Fig. 3(b)).

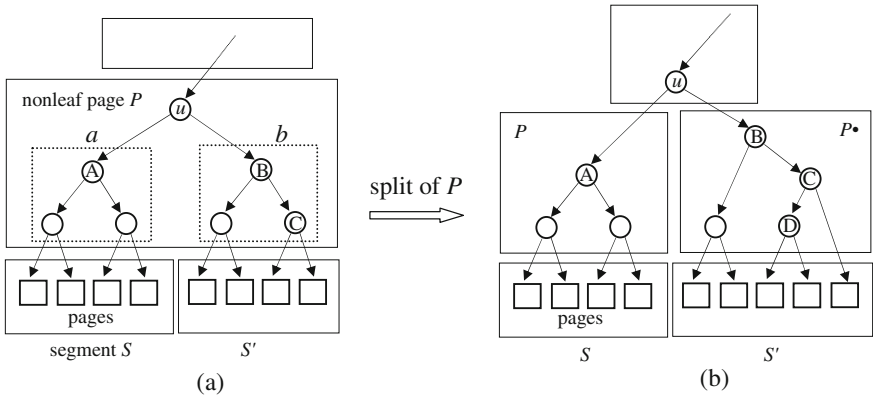


Fig. 3. Effect of a nonleaf page split

The SP-tree is a binary tree, and the entries in the left subtree and the right subtree of the local root entry in a certain nonleaf page point to the pages in different segments unless all the pages referenced lie in the same single segment. For example, the entries in the subtrees *a* and *b* in Fig. 3(a) point to the pages in different segments *S* and *S′*, respectively. When the page *P* is split, as shown in Fig. 3(b), the index entries that point to the pages in a certain segment are inserted into the same page. Thus, in the SP-tree, the references of pages that belong to a segment are stored in a single nonleaf page.

We should note that the SP-tree does not cause a split of lower level segment that is not full due to a split of upper nonleaf page. In the SP-tree, the fanout of a nonleaf page (it is often several hundreds) is far larger than that of a segment (it is usually several tens). In other words, there are generally many segments under a nonleaf index page. Therefore the segments referenced by the left and the right parts of the control node (i.e., local root) of the upper nonleaf page that is full are already different as shown in Fig. 3(a). Thus, after the upper level page split, it is not the case to split lower level segments to preserve the property that the references of pages that belong to a segment are stored in the same nonleaf page.

4 Performance Evaluation

To demonstrate the practical effectiveness of the SP-clustering, we implemented the SP-tree by using 64-KB segments. We performed an extensive experimental evaluation of the SP-tree and compared it to the pure LSD^h-tree. The experimental results demonstrate that in most cases the SP-clustering is superior to traditional MIMs. For random queries such as exact-match queries and nearest neighbor queries, there is little performance difference between the SP-tree and the LSD^h-tree.

5 Conclusions

We have introduced the SP-tree for range queries. The performance advantage of the SP-clustering comes from saving much disk startup time. Moreover, storing a sequence of index pages contiguously within a segment provides a compromise between optimal index node clustering and the excessive full index reorganization overhead. Thus, the SP-clustering methods may be used as an alternative index clustering scheme. The SP-clustering is so generic that it may be applied to most MIMs.

References

1. Henrich, A.: The LSDh-tree: an access structure for feature vectors. In: Proceedings of the International Conference on Data Engineering, pp. 362–369 (1998)
2. Henrich, A., Six, H.-W., Widmayer, P.: The LSD-tree: spatial access to multidimensional point and non-point objects. In: Proceedings of the ICDE, pp. 44–53 (1989)
3. Litwin, W., Lomet, D.B.: The bounded disorder access method. In: Proceedings of the IEEE International Conference on Data Engineering, pp. 38–48 (1986)
4. Lomet, D.B.: A simple bounded disorder file organization with good performance. ACM Trans. Database Syst. **13**(4), 525–551 (1988)
5. O'Neil, P.E.: The SB-tree: an index-sequential structure for high-performance sequential access. Acta Informatica **29**, 241–265 (1992)

An Address Conflict Resolving Scheme of Inter-drone Ad Hoc Communications for Hide Densely Deployed Low Power Wide Area Networks

Jaeho Lee[1] and Bong-Ki Son[2(✉)]

[1] Department of Information and Communications Engineering, Seowon University, 316, 1st Science Building, 377-3 Musimseoro, Seowon-Gu, Cheongju, Chungbuk 28674, South Korea
izeho@seowon.ac.kr
[2] Department of Computer Engineering, Seowon University, 315, 1st Science Building, 377-3 Musimseoro, Seowon-Gu, Cheongju, Chungbuk 28674, South Korea
bksohn@seowon.ac.kr

Abstract. Most of many communication technologies employed address identification method but many of them have not presented appropriate solution in which communication nodes would be widely and densely deployed. In this case, manually assigned ID generation method can be aggravately inefficient where there can have heterogeneously manufactured devices. In swarm flight of drone environmet, it can be high probability that multiple drones which were individually manufactured by different manufactural companies can configured with the same ID values, however, there is nothing solution in the current standard specifications of ad hoc communications, representatively as IEEE 802.15.4 or Zigbee standard. In order to find practical solution on the real world, we present an appropriate solution for dynamic ID generation with low conflict probability and for detection and avoidance method of ID conflict.

Keywords: Inter-drone communications · Ad hoc · Swarm flight · Address conflict · Dynamic ID generation

1 Introduction

Recently, the theme of Drone has been increasingly spotlighted under the unmanned system such as field discovery, measuring environmental parameters, disaster surveillance, and unmanned delivery service with another emerging theme of IoT(Internet of Things). Furthermore, swarm flight drone systems have been appearing as state of the art technology and this have been extremely promoting inter-drone communications.

In the actual environment, most of drones launched into the world market have just employed Wi-Fi as a role of video stream transmissions and FM analog radio for remote controlling drone attitude and position. However, for swarm flight of drone, communicational function between clustered drones should provide long range to cover

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_17

wide area and ad hoc delivery method with high energy efficiency. Hence, the realistic solution for above requirement can be in Zigbee technology which was designed for low-powered ad hoc communications based on MAC and PHY specifications defined from IEEE 802.15.4e and IEEE 802.15.4 g respectively.

On the other hand, many Zigbee devices recently launched on the real market provided only manual ID configuration method because the specifications of this technology cannot present dynamic or adaptive ID generation method, so every communication device has been just following with manufacture-compliant approach; most of all manufacturer employed 48bit IEEE address structure but didn't provide ID assignment method. If any operator employed heterogeneously manufactured drone products on the same application field, there is high probability of ID conflict on the same field.

For addressing this problem with the practical scenario, we designed a new solution composed of dynamic ID generation, based on Hash mechanism for reducing the probability of ID conflict, and effective detection and avoidance method for incurred ID conflict.

2 Dynamic PANID Generation and Conflict Avoidance

2.1 Hash Based Dynamic PANID

In Zigbee technology, the structure of all communication nodes was composed of PANID, for routing mechanism on each PC(Personal Coordinator) on widely deployed Ad hoc environment, and Node_ID, for local routing mechanism belonging to a PC; as a result, every PC just finds PANID for inter-PC routing. In other word, overall ad hoc routing function can be possibly miss-operated if any PC was configured with conflicted or duplicated ID value to any other PC on the whole networks. In general, IEEE address policy clearly gives unique ID function but it is difficult to be employed on MAC and routing layers due to the high length, so this cannot resolve the above problem because the specifications of IEEE 802.15.4e and IEEE 802.15.4 g employed only PANID which defined as 16bit length but not 48bit IEEE address.

This subsection presents dynamic PANID generation method based on Hash algorithm, as shown in Fig. 1. On this mechanism, every PC generates its PANID based on Hash algorithm which uses 48bit IEEE address of given PC as a seed value. So every PC has lowest probability of PANID duplication.

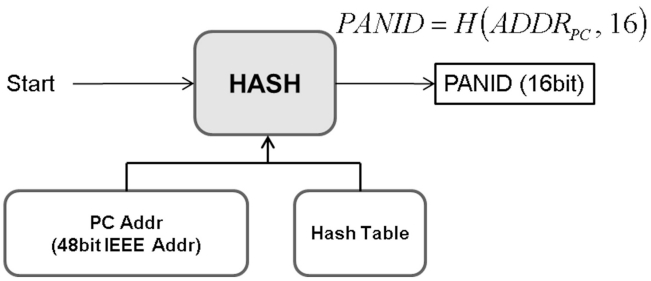


Fig. 1. Hash-based PANID generation method.

After a PC generated PANID with Hash, it should disseminate its PANID for reducing the probability of PANID confliction, because dynamic PANID generation scheme can make less conflict probability by considering neighbor's PANID if it can be aware. So this scheme makes initial flooding mechanism which let every PC sends dynamically generated PANID to its neighbor PC with broadcast transmission, as shown in Fig. 2.

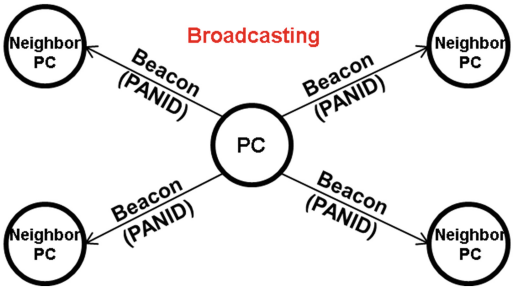


Fig. 2. PANID broadcast mechanism from the PC which just makes dynamic PANID.

If a PC detects confliction between own PANID and received PANID broadcasted from neighbor area, it can discard own PANID and generate PANID again with the different seed value.

2.2 PANID Conflict Detection and Avoidance Method

Even though the previous subsection illustrated how it provides lowest PANID confliction probability with Hash algorithm and unique 48bit IEEE address, there is still probability of PANID confliction when the network size grows up to large scaled swarm flight environment of drone. Hence, we additionally designed PANID conflict detection method and avoidance method, as shown in Figs. 3 and 4 respectively. In Fig. 3, the PC located on the middle of the figure can overhear every packet which would be delivered via the PC, so it can be aware of PANID confliction derived from far-located PC. In ad hoc environment, every packet would be delivered in hop by hop, so every intermediate PC can check the PANID included on the packet and check whether it is conflicted with its PANID or not.

Moreover, if a PC detects PANID confliction with long-distanced PC, the PC detecting the conflict would notify this event to the originated source PC by flooding the packet including the event of PANID confliction. The reason of using flooding method is that all PCs have to know the conflicted PANID on the whole ad hoc environment because they should change their PANID if any PC has the same PANID with the event packet.

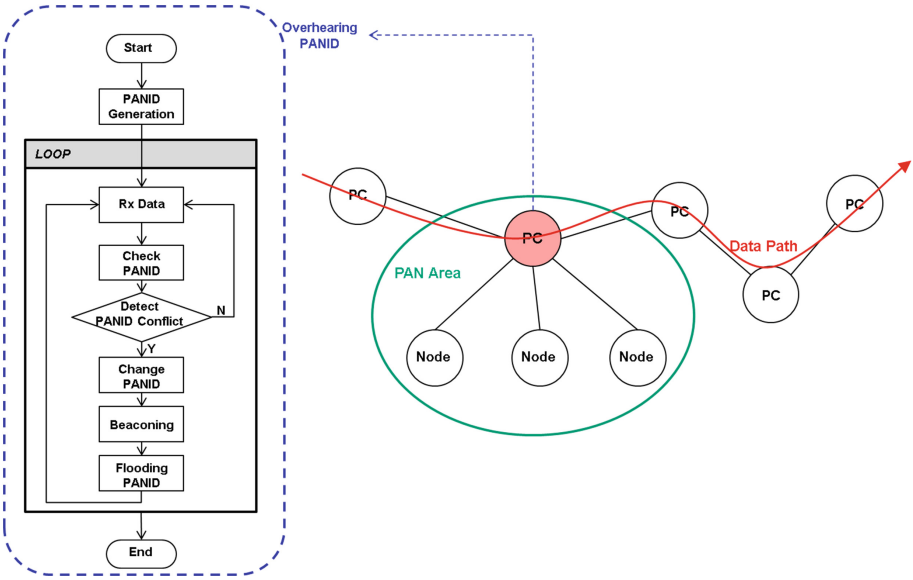


Fig. 3. PANID conflict detection method.

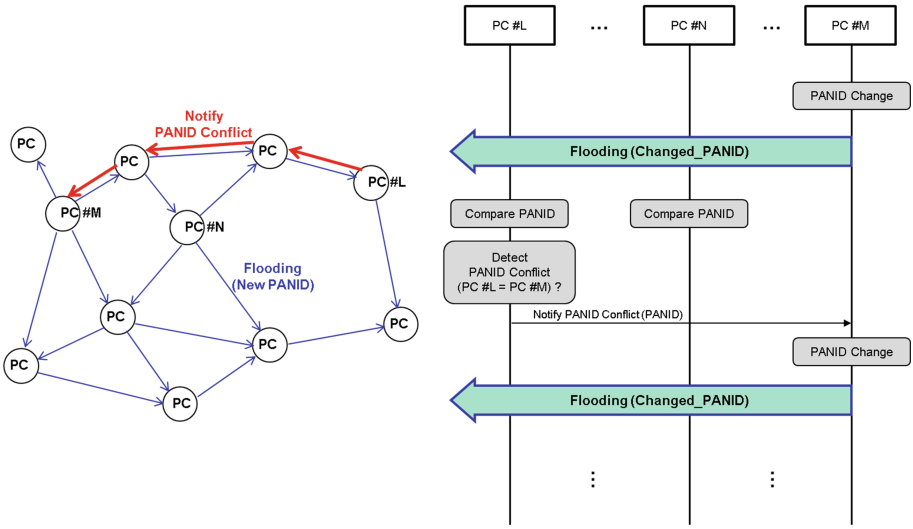


Fig. 4. Notification for PANID conflict.

3 Evaluation

Depending on the proposed scheme illustrated on the previous section, we performed simulation-based experiments to verify the performance of our scheme for the evaluations. Figure 5 shows the results of PANID confliction event number and the PDR (Packet Delivery Ratio) according to the traffic increment on the whole ad hoc environment. For the objective perspective, we employed DSDV as a representative proactive routing algorithm and AODV as a representative reactive routing algorithm. In these results, we found that proactive routing algorithm can present lower PANID conflict event comparing with reactive routing, and also found that PANID confliction could be overcome according to increment of the packet delivery frequency. Moreover, regardless of PANID confliction increment, we can found that PDR would be maintained up to 82 % on the whole traffic environment. With these results, we can verify the proposed scheme can basically prevent the confliction event of PANID and can also address from PANID confliction with the avoidance scheme we proposed.

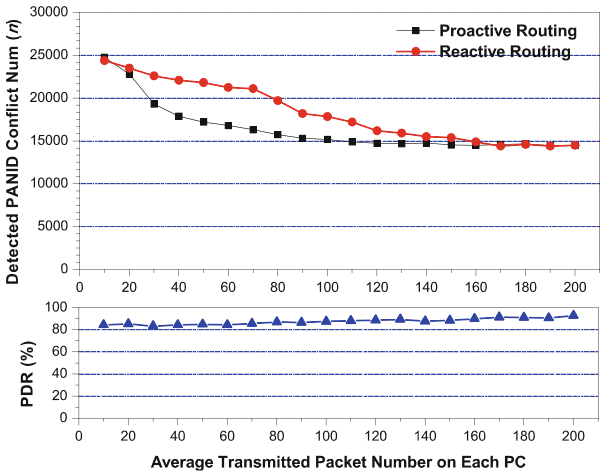


Fig. 5. Number of PANID confliction and PDR according to the traffic amount.

4 Conclusion

We presented dynamic PANID generation method based on Hash algorithm with the seed of 48bit unique IEEE address, and illustrated conflict detection and avoidance method, and prove the performance of the proposed scheme with the results of simulation-based experiments, for verifying that the proposed scheme will be useful on the environment of drone swarm flight which has the characteristics of high PANID conflict probability due to the heterogeneously manufactured multiple drones. For further research, we will design MAC protocol for meeting the fundamental requirement of massively deployed drone environment, with continuing this proposed scheme.

References

1. Zigbee Specification, Zigbee Alliance Inc., September 2012
2. http://www.ieee.org
3. IEEE, IEEE Standard for Local and Metropolitan Area Networks, Part 15.4 (Low-Rate Wireless Personal Area Networks), September 2011
4. Chiang, C-C., Wu, H-K., Liu, W., Gerla, M.: Routing in clustered multihop, mobile wireless networks with fading channel. In: Proceedings of IEEE SICON, pp. 197–211, April 1997
5. Perkins, C.E., Royer, E.M.: Ad-hoc on-demand distance vector routing. In: WMCSA 1999 Proceedings of the Second IEEE Workshop on Mobile Computing Systems and Applications, pp. 90–100, February 1999
6. Haas, Z.J., Pearlman, M.R., Samar, P.: The zone routing protocol (ZRP) for ad hoc networks. Internet Draft, draftietf-manet-zone-zrp-04 (2002)
7. Johnson, D.B., Maltz, D.A.: Dynamic source routing in ad-hoc wireless networks. In: Imielinski, T., Korth, H.F. (eds.) Mobile Computing. The Kluwer International Series in Engineering and Computer Science, vol. 353, pp. 153–181. Springer, Heidelberg (1996)
8. Lee, J.: A new routing scheme to reduce traffic in large scale mobile ad-hoc networks through selective on-demand method. Wirel. Netw. **20**(5), 1067–1083 (2014)
9. Wang, L., Olariu, S.: A two-zone hybrid routing protocol for mobile ad hoc networks. IEEE Trans. Parallel Distrib. Syst. **15**, 1105–1116 (2004)
10. Busch, C., et al.: Approximating congestion + dilation in networks via 'quality of routing' games. IEEE Trans. Comput. **61**(9), 1270–1283 (2012)
11. Lee, J.: A traffic-aware energy efficient scheme for WSN employing an adaptable wakeup period. Wirel. Pers. Commun. **71**(3), 1879–1914 (2013)
12. Saxena, N., Roy, A., Shin, J.: A QoS-based energy-aware MAC protocol for wireless multimedia sensor networks. In: Proceedings of Vehicular Technology Conference (VTC), pp. 183–187, May 2008
13. Lu, C., Blum, B., Abdelzaher, T., Stankovic, J., Tian, H.: RAP: a real-time communication architecture for large-scale wireless sensor networks. In: Proceedings of IEEE Real-time Systems Symposium (RTSS), pp. 55–66, December 2001
14. Lee, J.: A massive transmission scheme in contention-based MAC for wireless multimedia sensor networks. Wirel. Pers. Commun. **71**(3), 2079–2095 (2013)
15. Weniger, K.: PACMAN: passive autoconfiguration for mobile ad hoc networks. IEEE JSAC, Wirel. Ad Hoc Netw. **23**, 507–519 (2005)

State-of-the-Art Algorithms for Mining Up-to-Date High Average-Utility Patterns

Donggyu Kim and Unil Yun(✉)

Department of Computer Engineering, Sejong University, Seoul, Korea
donggyukim@sju.ac.kr, yunei@sejong.ac.kr

Abstract. High average-utility pattern mining is an emerging issue in the association rule mining area due to its meaningful mining results reflecting the characteristics of items such as their importance and quantities, and consideration on lengths of patterns. Recently, various studies have been dedicated to the researches of mining methods that extract up-to-date patterns from stream data, which are continually generated from various sources without limitations. A sliding window technique is one of methods for handling such stream data and mining up-to-date patterns. In this paper, we introduce state-of-the-art algorithms for finding up-to-date high average-utility patterns over data stream by using the sliding window method.

Keywords: Data mining · Utility pattern mining · High average-utility pattern mining · Stream pattern mining

1 Introduction

High average-utility pattern mining [4] was proposed to solve the problem of traditional high utility pattern mining [7] that utilities of patterns tend to be larger as lengths of them become longer. Since high average-utility pattern mining calculates utilities of patterns by the summations of item utilities in patterns divided by the corresponding pattern lengths, such negative tendency in high utility pattern mining can be eliminated from the mining procedure of high average-utility pattern mining. Meanwhile, mining up-to-date high average-utility patterns from stream data is one of challenging issues because massive and continuous stream data are generated from various sources in recent years. In this paper, we introduce the most recent high average-utility pattern mining algorithm finding up-to-date high average-utility patterns over data stream.

The reminder of the paper is organized as follows. In Sect. 2, the explanation on related works such as a sliding window technique is given. In Sect. 3, we introduce state-of-the-art high average-utility pattern mining algorithms and discuss the differences between algorithms for mining patterns from static databases and data streams. In Sect. 4, we evaluate the performances of algorithms using real world datasets. Finally, we conclude the paper in Sect. 5.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_18

2 Related Works

Sliding window techniques [6] have been employed in many pattern mining approaches such as frequent pattern mining, sequential pattern mining, and high utility pattern mining in order to find up-to-date interesting pattern information. A specific time period called *window* is used in sliding window based methods. *window* consists of multiple chunks of data called *batch*, respectively. The data belonging to the oldest batch is removed from *window* whenever new data are generated from a data stream. Then, the new data are inputted into *window* as the most recent batch. Therefore, only recent data can be maintained in *window*.

There are other methods for mining recent patterns from stream data. A damped window technique [2] assigns more weights to recent data by gradually decaying the importance of old data as time passes. In this method, a parameter called *decaying factor* is used to control the weight of old data. On the other hand, a landmark window technique [5] only considers data belonging to a specific time period, which is between the current time and a user-specified time point called *landmark*.

3 High Average-Utility Pattern Mining Algorithms Using Sliding Window

The first high average-utility pattern mining algorithm, TPAU [4] (*Two-phase high average-utility pattern mining*), has its root in Apriori algorithm [1]. It generates candidate patterns with (k+1)-lengths by joining candidate patterns with k-lengths. At each join operation, a database scan is required in order to determine whether patterns are valid candidate patterns or not. Therefore, it has many flaws such as exhaustive database scans and huge candidate pattern generation. In addition, since this algorithm focuses on the mining process for static databases and does not consider the generation time of data in its mining process, it cannot find up-to-date high average-utility patterns.

On the other hand, SHAU [8] (*Sliding window based high average-utility pattern mining*) is a state-of-the-art algorithm that employs a sliding window technique in order to find up-to-date patterns over data streams. The algorithm employs its tree structure in order to avoid constant database scanning. Its tree structure is similar with that of FP-Growth [3], which is a representative pattern mining algorithm using a tree structure. However, each node in its tree structure has an additional array structure in order to manage information of batches separately. Therefore, the algorithm can remove the information of the oldest batch from the tree structure efficiently. By scanning given stream data only one time, SHAU constructs its tree structure that stores all necessary

Table 1. Comparison of TPAU and SHAU

Algorithm	Data structure	Database scanning	Stream data processing	Mining approach
TPAU	None	Multiple times	No	Apriori approach [1]
SHAU	Tree structure	Only two times	Yes	Pattern growth approach [3]

information compactly. Table 1 shows the comparison of TPAU and SHAU. This method also reduces the number of candidate patterns by using its novel strategy called RUG (*Reducing upper-bound in global tree*). The strategy uses the concept of CMU (*Conditional maximal utility*) for minimizing overestimated average-utility values in its tree structure. Traditional algorithms [4] use the concept of a maximum utility as the upper-bound of an average-utility in a transaction. When $u(i, T)$ indicates the utility of item i in transaction $T = \{i_1, i_2, \ldots, i_k\}$, a maximum utility is defined as the following formula.

$$mu(T) = max(u(i_1, T), u(i_2, T), \ldots u(i_k, T)) \tag{1}$$

On the other hand, SHAU utilizes CMU as the upper-bound of an average-utility in a transaction instead of a maximum utility. CMU is defined as the following formula.

$$cmu(i_n, T) = max(u(i_1, T), u(i_2, T), \ldots u(i_n, T)) \, (1 \leq n \leq k) \tag{2}$$

As defined in the above formulas, since the value of CMU can become lower than the maximum utility according to the utilities of items consisting a transaction, SHAU can reduce the upper-bound values stored in its tree structure. Due to the utilization of such novel strategies, SHAU generates much smaller candidate patterns than previous algorithms.

4 Performance Analysis

In this section, we analyze the performance of algorithms in terms of their runtime and memory performances in stream data environments. Since a state-of-the-art algorithm, SHAU, is only sliding window based algorithm, we use the modified version of TPAU (denoted as STPAU) using sliding window for fairly comparing the performances of SHAU and existing algorithms. A PC with 3.30 GHz CPU, 8 GB RAM, and Windows 7 OS is employed in order to execute algorithms. A real dataset, Chess, is employed for experiments. Table 2 shows the characteristics of real datasets used for experiments. Each symbol signifies the following characteristic of a dataset. |D| is the total number of transactions in a dataset. |I| is the total number of distinct items in a dataset. Finally, |T| is the average length of transactions in a dataset. The execution times of algorithms are measured on the total runtime for conducting mining processes for all windows. The memory spaces consumed by algorithms are measured on the peak memory during their entire mining processes.

Figure 1 presents experimental results obtained from the tests using Chess. The tests are performed by decreasing minutil from 5.5 % to 5.1 %. On the other hand, the sizes of each batch and a window are fixed to 1000 and 3, respectively. We can observe

Table 2. Characteristics of dataset.

Dataset	|D|	|I|	|T|
Chess	3196	75	37

that SHAU has much better performance than STPAU in terms of runtime and memory consumption. In addition, we can see that the execution time of SHAU is gently increased as the threshold is decreased while that of TPAU is exponentially increased according to the decreased of the threshold. The reason is that the number of candidate patterns generated by STPAU is significantly increased as the threshold becomes lower, and huge execution time is required in order to validate candidate patterns. We can also see that SHAU consumes much less memory compared to STPAU regardless of minutil settings. The reason is that STPAU generates a lot of candidate patterns and uses huge memory space in order to load candidate patterns in main memory because STPAU needs to handle candidate patterns by calculating actual average-utilities of them. Even though SHAU has to construct its tree structure storing the information of a database, SHAU requires much smaller memory space compared to STPAU because it constructs its tree structure compactly through the node sharing effect [3] and the number of candidate patterns generated by SHAU is much smaller than that of STPAU.

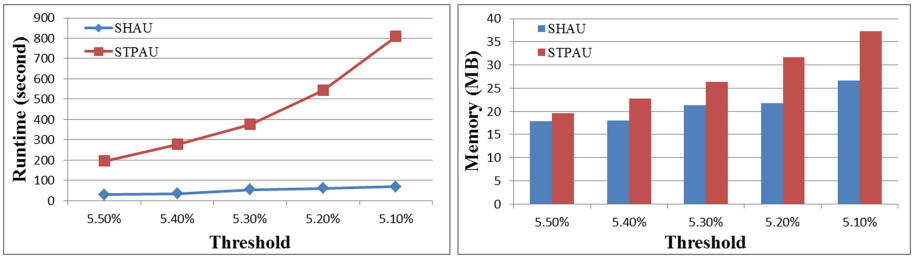


Fig. 1. Experimental results

5 Conclusion

In this paper, we introduced state-of-the-art sliding window based algorithms for mining up-to-date high average-utility patterns over data streams. In addition, we did the performance analysis by comparing the runtime and memory performances of algorithms. In order to efficiently process given stream data and mine valuable patterns from the stream data, we need fast algorithms that consider stream environments. In this regard, the state-of-the-art high average-utility pattern mining algorithms developed so far have fundamental limitations because they have to validate candidate patterns generated from their mining processes using an apriori approach or a pattern growth approach. Therefore, further studies on more efficient algorithms for mining up-to-date patterns over data streams should be conducted in future works. In addition, researches on methods using other stream pattern mining techniques such as damped window and landmark window also can be conducted for mining up-to-date high average-utility patterns efficiently.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 2015206 2051 and NRF No. 20155054624).

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
2. Chen, L., Mei, Q.: Mining frequent items in data stream using time fading model. Inf. Sci. **257**, 54–69 (2014)
3. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining Knowl. Discov. **8**(1), 53–87 (2004)
4. Hong, T.-P., Lee, C.-H., Wang, S.-L.: Effective utility mining with the measure of average utility. Expert Syst. Appl. **38**, 8259–8265 (2011)
5. Li, H., Shan, M., Lee, S.: DSM-FI: an efficient algorithm for mining frequent itemsets in data streams. Knowl. Inf. Syst. **17**(1), 1151–1163 (2008)
6. Ryang, H., Yun, U.: High utility pattern mining over data streams with sliding window technique. Expert Syst. Appl. **57**(15), 214–231 (2016)
7. Yun, U., Ryang, H.: Incremental high utility pattern mining with static and dynamic databases. Appl. Intell. **42**(2), 323–352 (2015)
8. Yun, U., Kim, D., Ryang, H., Lee, G., Lee, K.: Mining recent high average utility patterns based on sliding window from stream data. J. Intell. Fuzzy Syst. **30**, 3605–3617 (2016)

Design of Shoot'em up Game Using OpenGL

Unil Yun(✉) and Heungmo Ryang

Department of Computer Engineering, Sejong University, Seoul, Korea
yunei@sejong.ac.kr, ryangl@sju.ac.kr

Abstract. The OpenGL is an Open Graphics Library developed for rendering 2D and 3D vector graphics. This library has been widely used in various areas such as game development, simulation, and visualization. In this paper, we aim to design a shooting simulation program with three objects based on the OpenGL. For this purpose, we analyze collision detection techniques that are an essential element in shooting game development and suggest an appropriate method.

Keywords: Collision detection · Shooting game · OpenGL

1 Introduction

With the advent of computer graphics technology, the game industry has also grown rapidly. In addition, various game development engines such as unity [7], OpenGL [6], and unreal [8] have been emerged and helped developers to make their games easily, and after the emergence, related researches have been studied [2–4]. Among the engines, the OpenGL is an Open Graphics Library developed by the Silicon Graphics for rendering 2D and 3D vector graphics. Including the game development, this library has been widely used in many applications such as simulation and visualization. In this paper, we design a shooting simulation program based on the OpenGL.

Shooting games are one of game genres where players defeat enemies by shooting missiles and this genre has become popular. Figure 1 is a famous shooting game, Strikers 1945 developed by the Psikyo and released in 1995, where each player controls a plane and defeats enemies by attacking them with the plane's missiles and bombs. One of immersion factors of the shooting games is immediate reaction of the enemies when they are shot down by the players' missiles. For this purpose, collision detection [1, 5] is utilized, and this is one of essential elements in shooting game development and decides whether two or more objects are intersected. To make the shooting games more real, there is a need to understand characteristics of collision detection algorithms and provide the users with immediate and acceptable collision results by the collision detection algorithms. For the improvement of a basic detection algorithm, we suggest a hybrid detection approach in this paper.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_19

Fig. 1. Strikers 1945

2 Related Work

There are various collision detection algorithms [1, 5] such as circle collision detection, per-pixel collision detection that shows precise results in 2D spaces, simple rectangle collision detection, and rotated rectangle collision detection. Figure 2 is an example of the circle collision detection. In the left of the figure, no collision between two objects, O_A and O_B, is detected based on the distance between their central points, C_A and C_B. On the contrary, their collision occurs in the right of the figure. Although the simple rectangle collision detection approach can decide collisions fast, it considers object areas as rectangle shapes regardless of their real shapes. As a result, this approach may make a wrong decision when there is no visual collision, which can lower reality. Therefore, in this paper, we propose a hybrid collision detection method using rectangle and diamond shapes. By using the proposed algorithm, we can detect collisions fast like the rectangle one and improve accuracy with diamond areas.

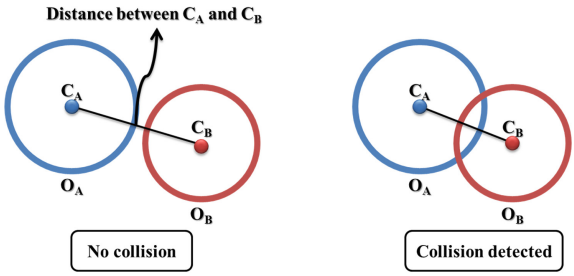


Fig. 2. Example of circle collision detection

3 Game Design

In this section, we explain and design main objects and our collision detection algorithm. There are three types of objects in our program: a cannon, its cannonball, and a flying object. Figure 3 shows a demo program with the objects. The cannon object is drawn with a circle and a long rectangle. This object rotates clockwise or anticlockwise according to a user's keyboard input. To implement the rotation using the OpenGL [6],

we have to know a coordinate value of the rotated object. For this purpose, a formula to calculate the base's length and the height of a right-angled triangle is used. Based on the computed values, we move the cannon. After the rotation, its cannonball can be fired in the direction the cannon points. The flying object moves from the right to the left using the timer function of the OpenGL. In here, the timer function is employed to move the flying object whenever a fixed period of time is passed. To improve simplicity of the flying object's appearance and movement, moreover, our program uses a random value to set the period, through which it makes prediction for the target difficult. In the program, users have to attack the flying object by adjusting the direction of the cannon and firing its cannonball.

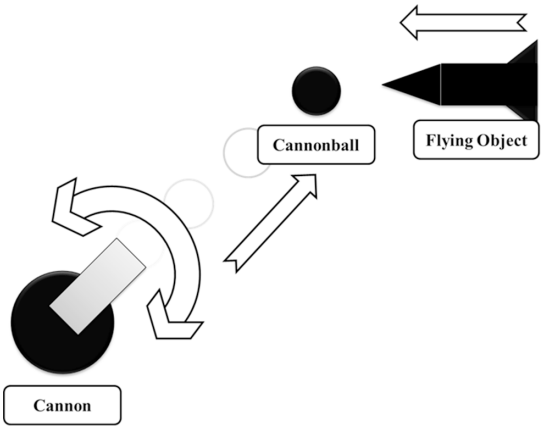


Fig. 3. Demo program with three types of objects

In our program, a collision detection algorithm using rectangle and diamond shapes is utilized to decide whether a cannonball is intersected with the flying object or not. Assume that there are two cross-shaped objects, O_A and O_B, and O_B is approaching O_A as shown in Fig. 4(a) and (b).

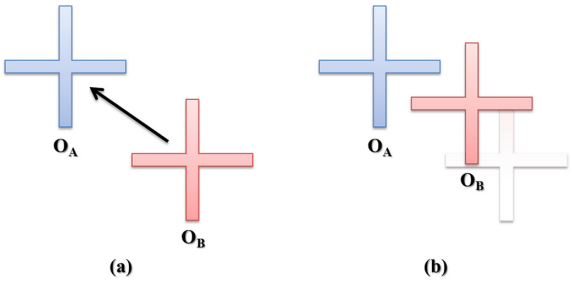


Fig. 4. Two cross-shaped objects

In case of the approach, a basic rectangle method can detect a wrong collision as shown in Fig. 5(a). This issue can be improved with our hybrid method in Fig. 5(b). That is, our algorithm chooses one to detect a collision between two objects according to a situation.

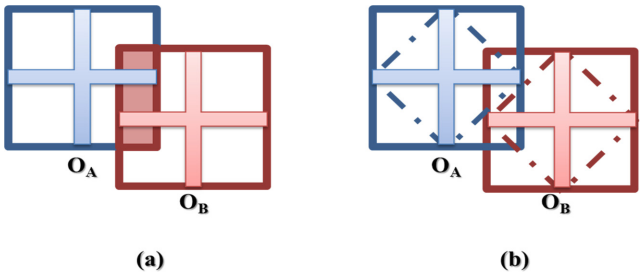


Fig. 5. Collision detection with rectangle and diamond shapes

Let the rectangle areas of O_A and O_B be R_A and R_B and their diamond areas be D_A and D_B in Fig. 5. Then, the proposed algorithm conducts its collision detection with the following four conditions: (1) a collision of R_A and R_B, (2) that of R_A and D_B, (3) that of D_A and R_B, and (4) that of D_A and D_B. Only when all of the conditions are satisfied, our algorithm decides a collision between the two objects. For O_A and O_B in Fig. 5(b), for example, intersections between R_A and R_B, R_A and D_B, and D_A and R_B occur, but there is no intersection between D_A and D_B. Hence, the algorithm decides that there is no collision between them. In real situations, however, objects are unlikely to have cross shapes (extreme case) as shown in Fig. 5. For this reason, we apply the method with the four conditions to the program when the percent of one object's area inscribed in a rectangle area of the other is smaller than 50 %. The reason for this is that the rate of inscribed diamond area is 50 %. In shooting games, since immediate and fast reactions are necessary, the proposed algorithm has an advantage compared to the simple rectangle method in that it conducts collision detection with similar amount of operations.

4 Conclusion

In this paper, we designed a shooting simulation program developed using the OpenGL and proposed a hybrid collision detection algorithm using rectangle and diamond shape. The proposed algorithm could improve detection performance of a basic rectangle detection algorithm.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624).

References

1. Guo, K., Xia, J.: An improved algorithm of collision detection in 2D Grapple Games. In: Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 328–331, Jinggangshan (2010)
2. Jiyuan, L., Wenfeng, H.: Development of puzzle game about children's etiquette based on Unity3D. In: 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, pp. 495–500, Shanghai (2011)
3. Lewis, J., Brown, D.J., Cranton, W., Mason, R.: Simulating visual impairments using the Unreal Engine 3 game engine. In: 2011 IEEE 1st International Conference on Serious Games and Applications for Health, pp. 1–8, Braga (2011)
4. Lin, Q., Zhao, Z., Xu, D., Wang, R.: Design and implementation of an OpenGL based 3D first person shooting game. Trans. Edutainment **5**, 50–61 (2011)
5. Ng, K.W., Yeap, Y.W., Tan, Y.H, Ghauth, K.I.: Collision detection optimization on mobile device for shoot'em up game. In: 2012 International Conference on Computer & Information Science, pp. 464–468, Kuala Lumpur (2009)
6. OpenGL Game Engine. https://www.opengl.org/
7. Unity Game Engine. http://unity3d.com
8. Unreal Game Engine. https://www.unrealengine.com/

Performance Analysis of Tree-Based Algorithms for Incremental High Utility Pattern Mining

Heungmo Ryang and Unil Yun(✉)

Department of Computer Engineering, Sejong University, Seoul, Korea
ryang@sju.ac.kr, yunei@sejong.ac.kr

Abstract. To overcome drawbacks of traditional pattern mining such as difficulty reflecting characteristics of real-world databases to pattern mining, high utility pattern mining has been proposed and researched. Since database sizes become larger incrementally in many real-world applications, there is a need of appropriate methods to deal with such databases for discovering useful information from them efficiently. For this purpose, various approaches have been suggested. In this paper, we compare and analyze algorithms for high utility pattern mining from dynamic databases by considering characteristics of incremental databases and utilizing tree-based data structures. Moreover, we study their characteristics and direction of improvements based on experimental results of performance evaluation.

Keywords: Data mining · Pattern mining · Utility mining · Incremental mining · High utility patterns · Tree-Based algorithms

1 Introduction

Pattern mining is one of data mining techniques to discover useful information from huge databases, and it extracts such information in pattern forms. Although traditional pattern mining [1, 3] has played an significant role in the data mining field, this approach not only treats all items in databases with the same importance but also represents item occurrence as a binary form, i.e., 0 or 1. In addition, it cannot reflect characteristics of real-world databases to pattern mining completely. High utility pattern mining has been proposed and researched to overcome the limitations. Furthermore, database sizes become larger incrementally in various real-world applications and previous general static methods are not suitable for finding meaningful information efficiently from such databases. Incremental high utility pattern mining [2, 4, 6] has been studied to mine essential information from dynamic databases by considering real characteristics of them.

The remainder of this paper is organized as follows. In Sect. 2, we describe influential researches related to tree-based incremental high utility pattern mining. In Sect. 3, we analyze characteristics of recent tree-based methods for incremental high utility pattern mining through experiments for performance evaluation with real datasets and study direction of improvements based on experimental results. Lastly, in Sect. 4, we summarize contributions of this paper.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_20

2 Related Work

In this section, we describe two types of related studies: traditional frequent pattern mining and high utility pattern mining for static databases.

2.1 Frequent Pattern Mining

Apriori [1] and FP-Growth [3] are well-known BFS and DFS frequent pattern mining algorithms respectively. The former utilizes a candidate generation-and-test approach, which causes the performance degradation since this method makes a large number of candidates and requires multiple database scans. The latter conducts a series of mining processes through only two database scans without candidate generation by employing a divide-and-conquer manner. In general, FP-Growth-like algorithms outperform Apriori-based ones. In pattern mining, meanwhile, the anti-monotone property [1] is used for efficient mining. This suggests that if a pattern is not valid due to its smaller frequency than a given threshold, then its super patterns are not also valid. When an invalid pattern is found in mining processes, the pattern's search space is can be eliminated, which improves mining efficiency. Therefore, satisfying this property is an essential criterion in pattern mining.

2.2 High Utility Pattern Mining

In utility mining, maintaining the anti-monotone property is not easy and Two-Phase [5] is the first algorithm that satisfies the property by applying the overestimation concept, called transaction weighted utilization (twu), to mining processes. This model generates candidates with no smaller overestimation values than a given threshold in the first phase and then identifies actual high utility patterns by computing their utility values through an additional database scan. Since the twu concept was suggested, although various static algorithms with the overestimation model have been developed, they are not appropriate for handling dynamic databases.

3 Tree-Based Incremental High Utility Pattern Mining

In contrast to the previous static approaches where the whole mining processes have to be conducted whenever a target database is increased, incremental high utility pattern mining methods simply reflect new information to the current data structures for dealing with the dynamic database. In incremental high utility pattern mining, algorithms build global data structures with the original databases, generate candidate patterns, and identify actual high utility patterns from the candidates. After that, when new transaction information is added to the original databases, the incremental approaches update the data structures, optimize them, and conduct mining processes. Figure 1 shows a simple overall process of incremental high utility pattern mining. In the figure, there are two types of data, the original and incremented ones. As explained

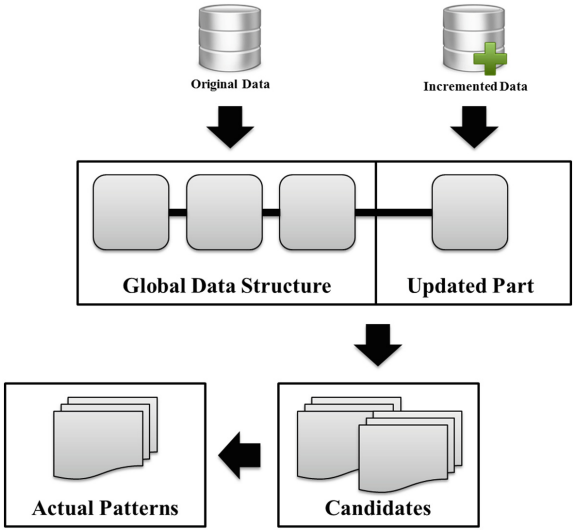


Fig. 1. Simple overall process of incremental high utility pattern mining

above, a set of high utility patterns is mined from the original data and then updated pattern results are extracted by handling only the incremented data.

FUP-HU [4] is an Apriori-based incremental high utility pattern mining algorithm and can reflect characteristics of real-world databases to pattern mining. However, this method requires a lot of database scans and operations. To address this issue, FP-Growth-based algorithms with tree data structures, IHUP [2] and HUPID [6], were proposed. IHUP applies the overestimation model of the Two-Phase algorithm [5] to its mining processes. For incremental high utility pattern mining, IHUP firstly constructs a tree data structure through a single database scan, extracts candidate patterns, and identifies actual high utility patterns from them with an additional scan. Although this algorithm mines valid patterns faster than the previous Apriori-based ones by utilizing the FP-Growth approach, it still generates a large number of candidates and consumes high computational time for the identification of actual patterns. To address this issue, HUPID extracts the fewer number of candidate patterns by reducing the overestimation values, through which it improves mining performance of tree-based incremental high utility pattern mining.

4 Performance Analysis

In this section, we evaluate mining performance of tree-based incremental high utility pattern mining algorithms, IHUP [2] and HUPID [6], in terms of runtime and the number of generated candidate patterns with the Chain-store dataset (http://cucis.ece.northwestern.edu/projects/DMS/MineBench.html), where real utility information is included. All performance experiments were conducted on an experimental environment with a 4.0 GHz Intel processor and 32 GB memory, running the 64bit Windows 7 OS.

For the performance evaluation in an incremental environment, we first used only 20 % of the dataset, and then added each 20 % of the rest to the initial data. With the incremental data, we repeatedly constructed tree data structures of the algorithms, restructured them, and conducted mining processes.

Figure 2 shows experimental results of the performance evaluation for IHUP and HUPID with the incremental Chain-store dataset in terms of runtime and the number of candidates when a minimum utility threshold is 0.03 %. From the figure, we can observe that runtime is proportional to the number of extracted candidate patterns. In the results, moreover, HUPID mines the same number of high utility patterns faster than IHUP by generating the smaller number of candidates.

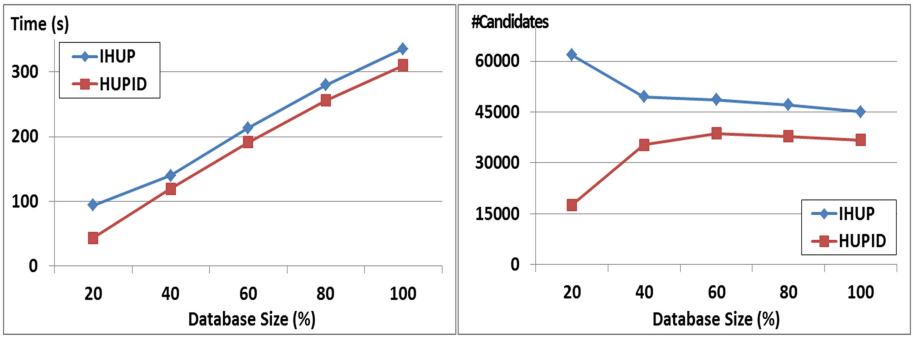


Fig. 2. Experimental results on chain-store

From the above experimental results, we can learn that performance of incremental high utility pattern mining algorithms depends on how many candidate patterns are extracted, and as a result developing an effective method for decreasing their number can be a good future work.

5 Conclusion

This paper studied characteristics of tree-based incremental high utility pattern mining algorithms and direction of their improvements based on experimental results for performance evaluation. As a result, we could learn that mining performance of an algorithm that generated the smaller number of candidates outperformed the other one and decreasing the number of extracted candidates can be a significant factor to improve performance of incremental high utility pattern mining.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624).

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th International Conference on Very Large Data Bases, Santiago de Chile, pp. 487–499 (1994)
2. Ahmed, C.F., Tanbeer, S.K., Jeong, B.-S., Lee, Y.-K.: Efficient tree structures for high utility pattern mining in incremental databases. IEEE Trans. Knowl. Data Eng. **21**(12), 1708–1721 (2009)
3. Han, J., Pei, J., Yin, Y.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. iData Min. Knowl. Disc. **8**(1), 53–87 (2004)
4. Lin, C.-W., Lan, G.-C., Hong, T.-P.: An incremental mining algorithm for high utility itemsets. Expert Syst. Appl. **39**(8), 7173–7180 (2012)
5. Liu, Y., Liao, W.-K., Choudhary, A.N.: A two-phase algorithm for fast discovery of high utility itemsets. In: Advances in Knowledge Discovery and Data Mining, Hanoi, pp. 689–695 (2005)
6. Yun, U., Ryang, H.: Incremental high utility pattern mining with static and dynamic databases. Appl. Intell. **42**(2), 323–352 (2015)

Development of 2D Side-Scrolling Running Game Using the Unity 3D Game Engine

Wooseong Jeong and Unil Yun(✉)

Department of Computer Engineering, Sejong University, Seoul, Korea
wsjeong@sju.ac.kr, yunei@sejong.ac.kr

Abstract. The Unity 3D game engine, which can develop games on most platforms such as IOS, Android, and web browsers, is known for a tool easier to use than other engines. In this paper, we develop a game where a sheep runs away from the threat of a rancher on the basis of the Unity game engine. The game is developed in order that a player feels thrill and tension by manipulating the sheep to pass obstacles, avoid projectiles, and run away from the chasing owner.

Keywords: Unity 3D engine · Game development · Side-Scrolling running game

1 Introduction

Recent game development has become easily accessible to non-specialists because of a variety of relevant development tools. Especially the Unity engine provides functions for making games, such as rigid body components for 3D physics and 2D physics, animation systems, Lighting and Rendering references, audio, Virtual Reality references, and a multi-scene editing. If users employ the above functions, they can more easily develop games based on C# and JavaScript. In this paper, we describe how to create games easily such tools even though we are beginners. In addition, we develop a 2D side-scrolling running game, named *Revolt of the sheep*, using the Unity game engine.

2 Related Work

Previous game development was mainly a way to create the overall system of game. In the past, game developers had to make games by themselves using the C language [10], various platforms [7], and Html5 [8]. In recent development approaches, difficulty has been lowered because of game engines. Therefore, beginners are also able to easily make games. Thus, many games were developed using the Unity engine [3, 4, 6]. There are various game engines widely used in these days such as the Unreal engine [1], the Cry engine [9], and the Unity engine. Among the above engines, we use the Unity engine because it is light and easy. We also need to select a proper genre in order to create a game. We need to choose the genre that even beginners can easily create. There are various game genres released so far such as: 2D, 3D [2], RPG, shooting [5], running, and arcade. Although the Unity engine is based on the 3D game development,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_21

it is also possible to develop a 2D game. This paper focuses on developing a 2D game in order to show how to use the tool easily.

3 Game Development

In this section, we introduce the implementation of our game and explain its scenario using Fig. 1. We also provide a table that describes the components used in the game to explain how the game is made briefly. Finally, we present a flow chart for the game.

(a) Game start scene (b) Stage scene (c) Game clear scene

Fig. 1. Game screens on web browser

A story of the game, *Revolt of the sheep*, says that *sheep* escapes from a ranch by passing obstacles and running away from the chasing owner. The game is deigned to feel the thrill and tension from projectiles, obstacles, and the owner. The game consists of five stages and their difficulty rises every time stages are switched. Figure 1 shows the web browser based game developed using the Unity engine. Figure 1(a) is a scene that is shown before players start the game. Figure 1(b) is a playing scene. Figure 1(c) is shown when the game is cleared.

The game scenario is as follows. If the *sheep*, a main character of the game, jumps over the *fence* at the end of each stage, the stage is cleared. If every stage is passed, the game is cleared. The passage of time is expressed by darkening the background increasingly whenever a stage is cleared. Then, the *sheep* escapes from a ranch when the last stage is cleared.

The components used in the game are explained in Table 1. Among them, moving objects are *sheep*, *rancher*, *arrow*, and *rockfall*. Meanwhile, fixed objects are *clear box*, *fence*, *obstacle box*, *thorn*, and *fire*. All of the objects do not crash into each other except for *sheep*. If *sheep* touches *clear box*, the stage is cleared. If *sheep* touches *fence* or *obstacle box*, it cannot move forward. Finally, if *sheep* touches the others, the stage is failed and restarted. A collision decision of each object is verified by using the 'Colliders' and 'Is Triggers' functions of the Unity engine. Dropping *rockfall* and jumping *sheep* are developed by using the 'Use Gravity' function. If a direction key is pressed, *sheep*'s speed is accelerated from zero to a speed limit. Otherwise, the speed is deaccelerated to zero. All of the above processes are developed using the Unity engine's functions and C#.

The overall process of the game is briefly diagrammatized in the flow chart format shown in Fig. 2. All of the screens in the game consist of scenes. If a specific event occurs, the scenes are switched another scene.

Table 1. Images and descriptions of the components in the game

Image	Name	Explanation
	Sheep	*Sheep* can be moved by arrow keys and a space bar.
	Rancher	*Rancher* chases *sheep* at a constant speed (slower than the speed of *sheep*).
	Clear Box	If *sheep* crashes into *clear box*, the stage is cleared.
	Life Bar	This bar displays the remaining life of the player. The life is reduced by 1 if the player fails the stage. If the life is 0, the game is over.
	Fence	*Fence* is an obstacle placed directly in front of *clear box* at the end of the stage.
	Obstacle Box	*Obstacle box* delays the speed of *sheep* because the sheep should jump it over.
	Thorn	If *sheep* touches *thorn*, the stage is failed.
	Fire	If *sheep* touches *fire*, the stage is failed.
	Arrow	*Arrow* flies out to the face of *sheep* at a constant speed. If *sheep* is hit by *arrow*, the stage is failed.
	Rockfall	*Rockfall* is dropped by gravity from the sky. If *sheep* is hit by *rockfall*, the stage is failed.

The process of the flow chart is as fallows. First, if a player presses a button at the game start scene, the scene is turned into the scene of stage 1. If *sheep*, which is controlled by arrow keys, touches the specific object in the stage scene, the collision decision occurs. At the time, the collision decision is classified into two types (*clear box* or the other objects).

If a collision occurs on *clear box*, the stage is cleared and the next step is started. If the stage is the last stage, the game is cleared. Otherwise, the stage scene is changed to the other one and the stage starts. If the game is cleared, the scene is turned into the game clear scene and the player can go to the game start scene by pressing a button.

On the other hand, if a collision occurs on the other objects, the player's life is reduced by 1. If the life is 0, the player returns to the game restart scene. If it is not 0, the player challenges the current stage again.

In the future work, we are scheduled to enhance the functions of the proposed program. For example, the functions can include: (1) pausing the stage, (2) giving up

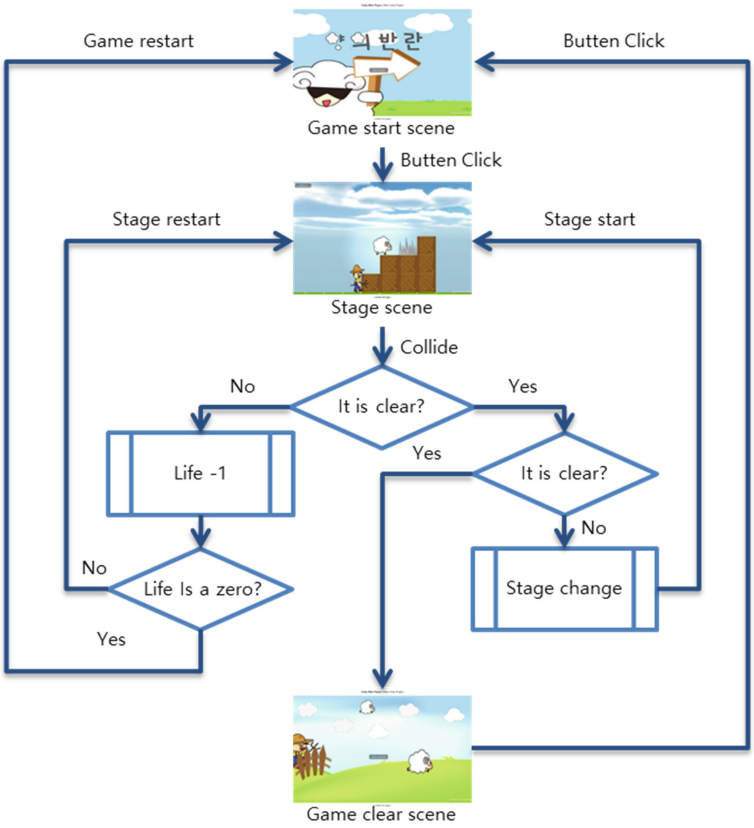


Fig. 2. The flow chart for *Revolt of the sheep*

the game, (3) using a variety of skills or items, and (4) increasing the vividness of the game by using the 'Animation System' of the Unity engine.

4 Conclusion

It is hard for beginners to develop games because they have to learn a variety of technics, skills, and knowledge for creating games successfully, not to mention they are time-consuming works. In this paper, we developed a game using the Unity engine that they can easily handle. Therefore, using the Unity engine is expected to lower the barriers to entry of game development.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624)

References

1. Bae, S.J., Kang, M.J.: Design and development of game AI using unreal engine 4 behavior tree. In: Proceedings of the Korean Society of Computer Information Conference, vol. 24(1), pp. 267–269 (2016)
2. Cuccurullo, S., Francese, R., Passero, I., Tortora, G.: A 3D serious city building game on waste disposal. Int. J. Dist. Educ. Technol. **11**, 112–135 (2013)
3. Dickson, P.E.: Using unity to teach game development: when you've never written a game. In: The 2015 ACM Conference on Innovation and Technology in Computer Science Education, Vilnius, pp. 75–80 (2015)
4. Kim, S.L., Suk, H., Kang, J.H., Jung, J.M., Laine, T.H., Westlin, J.M.: Using UNITY 3D to facilitate mobile augmented reality game development. In: IEEE World Forum on Internet of Things, Seoul, pp. 21–26 (2014)
5. Lin, Q., Zhao, Z., Xu, D., Wang, R.: Design and implementation of an OpenGL based 3D first person shooting game. In: Pan, Z., Cheok, A.D., Müller, W., Yang, X. (eds.) Transactions on Edutainment V. LNCS, vol. 6530, pp. 50–61. Springer, Heidelberg (2011). doi:10.1007/978-3-642-18452-9_3
6. Luo, M., Claypool, M.: Uniquitous: implementation and evaluation of a cloud-based game system in unity. In: Games Entertainment Media Conference, Toronto, pp. 1–6 (2015)
7. Narsoo, J., Sunhaloo, M.S., Thomas, R.: The application of design patterns to develop games for mobile devices using Java 2 micro edition. J. Object Technol. **8**, 153–175 (2009)
8. Parker, J.P.: HTML5 game development. In: CHI Conference on Human Factors in Computing Systems, Toronto, pp. 1011–1012 (2014)
9. Sherstyuk, A., Treskunov, A.: Head tracking for 3D games: technology evaluation using CryENGINE2 and faceAPI. In: IEEE Virtual Reality, pp. 67–68 (2013)
10. Zhang, W., Han, D., Kunz, T., Hansen, K.M.: Mobile game development: object-orientation or not. In: 31st Annual International Computer Software and Applications Conference, Beijing, pp. 601–608 (2007)

EPD Noticeboard for Posting Multiple Information

Bong-Ki Son[1] and Jaeho Lee[2(✉)]

[1] Department of Computer Engineering, Seowon University,
315, 1st Science Building, 377-3 Musimseoro, Seowon-gu,
Cheongju, Chungbuk 28674, Korea
bksohn@seowon.ac.kr
[2] Department of Information and Communications Engineering,
Seowon University, 316, 1st Science Building, 377-3 Musimseoro, Seowon-gu,
Cheongju, Chungbuk 28674, Korea
izeho@seowon.ac.kr

Abstract. In this paper, we propose an EPD noticeboard for effectively managing notice objects from various users and minimizing power consumption. The system posts more important and newer information based on authority level of the user, posting term and priority of the notice object. Because noticeboard with EPD requires power supply when data is updated, the power consumption is relatively low compared to LCD or OLED based systems. We also present one example implementation for managing digitalized notice objects in Korean university. E-paper display is quite suitable to be used for applications requiring low power and mainly using image data. We expect that our study shall be applied to mobile billboards, label, post-it and so on.

Keywords: EPD · Noticeboard · Notice object · Digital Poster · E-paper

1 Introduction

Important undergraduate schedule, employment information, announcement are commonly posted on the website of the university and also pinned to the noticeboard in each department. The assistant manually manages the noticeboard by receiving, classifying and putting on official paper, department notice, promotional material and so on. The important notice materials might be covered by others and out-of-date notices be posted due to limited space of the noticeboard, manual management and no prioritization.

The EPD (Electrophoretic Display) is used for showing visible images by vertically positioning charged pigment particles in microcapsules with the help of an external electrical filed. One of the most obvious advantages of the EPD is that it can provide a bi-stable display. Zero power is consumed to maintain the contents being displayed and the displayed image can be kept for a very long time after the removal of external power. Power will only be needed when one needs to the displayed contents and the power consumption will quickly go up if very frequent updates are required [1–3]. Thus EPD is quite suitable for applications which low power consumption is important

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_22

issue and image data is mainly used [4, 5]. EPD, also called E-paper is used on e-readers, smart watches, mobile phones, tag prices, etc.

In this paper, we propose an EPD noticeboard for effectively managing notice objects from various users and minimizing power consumption. The system posts more important and newer information based on authority level of the user, posting term and priority of the notice object. We also present one example implementation for managing digitalized notice objects in Korean university.

The remainder of this paper is organized as follows. Section 2 presents system architecture and main functions of subsystems in the proposed system. Section 3 describes the implementation and result. Finally, we summarize and conclude this paper in Sect. 4.

2 Proposed EPD Noticeboard

2.1 System Configuration and Architecture

Figure 1 shows configuration and architecture of the proposed system. Posting Client transmits information of the Digital Poster and JPG file translated document to post into Posting Server. Notice object images for the Digital Poster are merged into one image based on authority level of user, posting term and priority of the notice object. The merged notice object image is converted into 4-Grayscale and transmitted to the Digital Poster through Wi-Fi wireless LAN.

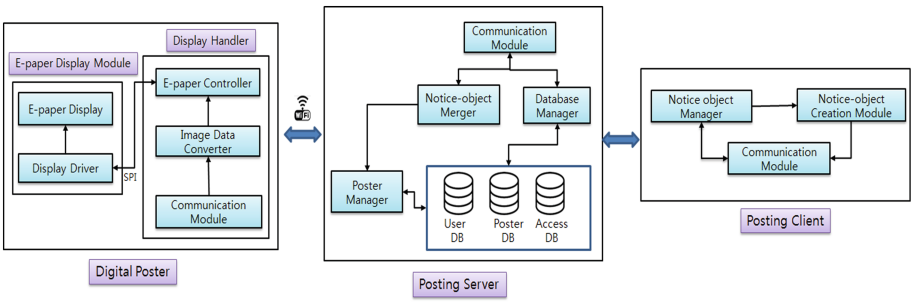


Fig. 1. System configuration and architecture.

4-Grayscale image is resized to the Digital Poster resolution and converted to E-paper data format. Finally, the notice object image is displayed in EPD by SPI (Serial Peripheral Interface) communication. Figure 2 shows process of displaying notice objects selected by users in our system.

2.2 Posting Server

When various users want to post notice object a Digital Poster, posting score N_i is computed through the following equation.

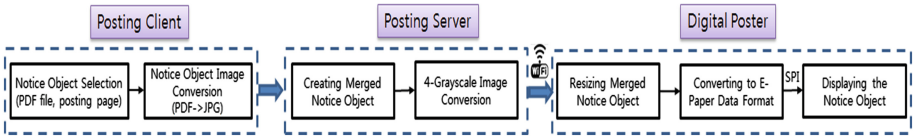


Fig. 2. The process of displaying notice objects.

$$N_i = A_i \times T_i \times P_i \tag{1}$$

where A_i is weight of user authority level about the Digital Poster and T_i is weight of posting term. P_i is weight of the notice object priority. Maximum number of notice object to be posted in a Digital Poster is limited. Thus notice objects in higher score are merged into one image.

Table 1 represents the factors for computing posting score to merge notice objects and its weights. Table 2 represents example of posting score in a Digital Poster which is able to post 4 notice objects. Therefore, notice object 10333, 10212, 10427, 10245 are merged into one image and posted in the Digital Poster. The notice object, 10111 is to be waiting state until posting term of a notice object is expired.

Table 1. The factors in computing posting score.

Factor	Weight
Authority level(A)	Super manager(3), Main manager(2), Accessor(1)
Posting term(T)	Within(3), Before(1), After(0)
Notice object priority (P)	Very high(3), High(2), Low(1)

Table 2. The example of posting score in a Digital Poster.

Notice object index	Authority level	Posting term	Notice object priority	Posting score
10333	Super manager (3)	Within (3)	Very high (3)	27
10212	Accessor (1)	Within (3)	Very high (3)	9
10427	Main manager (2)	Before (1)	High (2)	4
10245	Super manager (3)	Before (1)	Low (1)	3
10111	Super manager (3)	After (0)	High (2)	0

We use 4-Grayscale EPD in the proposed system. So, the merged color image is to be converted 4-Grayscale image in Posting Server and transmitted to the Digital Poster by Wi-Fi wireless LAN.

2.3 Digital Poster

Digital Poster resizes 4-Grayscale image for its display resolution and converts to image data format which 8-bits pixel is compressed into 2-bits pixel for processing in

the Display Driver. E-paper Controller transmits the compressed 4-Grayscale image into Display Driver by SPI. Finally, the Display Driver displays notice object image to E-paper Display, 4.3 in. SVGA EPD.

Figure 3 represents the sequence diagram for displaying notice object by SPI communication. Table 3 explains the functions of E-paper Controller and Display Driver for SPI communication.

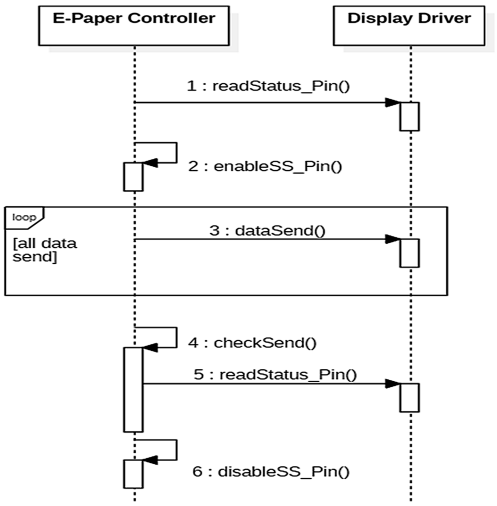


Fig. 3. Sequence diagram for displaying notice object by SPI communication.

Table 3. Functions of E-paper Controller and Display Driver for SPI communication.

Function name	Explanation
readStatus_Pin()	Check out that E-paper Display Driver is able to receive data by checking Status Pin
enableSS_Pin()	Activate SS_PIN of E-paper Display Driver
dataSend()	Transmit notice object image and command of clearing E-paper Display, displaying notice object image, etc.
checkSend()	Check out that notice object image and command are transmitted
disableSS_Pin()	Terminate communication by deactivating SS_PIN of E-paper Driver

3 Implementation and Result

Table 4 represents S/W environment for implementing the proposed system and Table 5 is H/W configuration of Digital Poster. Display Handler is executed on pcDuino3. Trial kit of LinkSprite [6] which consists of 4.3 in. SVGA EPD and E-paper Shild for Arduino is used as E-paper Display Module. Figure 4 shows screenshots of the H/W components.

Table 4. S/W implementation environment.

Role	OS	Implementation language	Etc.
Posting Server	Windows 8	Java	MySQL
Posting Client	Windows 8	Java	N/A
Digital Poster (Display Handler)	Linux 12.04	C/C ++	N/A

Table 5. H/W configuration of Digital Poster.

Device type		H/W	Specifications	
Display Handler		pcDuino3	CPU	AllWinner A20 SoC, 1 GHz ARM Cortex A7 Dual Core
			Onboard storage	4 GB Flesh Memory
			OS	Linux 12.04
			Network interface	Built-in WiFi, Ethernet 10 M/100Mbps
			Extended interface	Arduino sockets, 14xGPIO, 2xPWM, 6xADC, 1xUART, 1xSPI, 1xI2C
E-paper Display Module	E-paper Display	4.3″SVGA EPD (GDE043A2)	Active Matrix Electrophoretic Display, 800*600(SVGA) Display, White Reflectance above 35 %, 4:3 aspect ratio, 230 dpi	
	Display Driver	E-paper Shild for Arduino (ATmega32-16AC)	Atmel 8-bit Microcontroller, 1 MHz corespeed, 32 Kbyte Programmable Flash program memory, 1024 Bytes EEPROM, 2 Kbyte Internal SRAM, SPI, UART/USART extended, etc.	

(a) pcDuino3 Board (b)4.3inch EPD (c) Display Driver

Fig. 4. Configuration of Digital Poster.

Figure 5 shows the implementation results which Fig. 5(a) is sending notice object in Posting Client and Fig. 5(b) is to be posted 4 notice objects in higher posting score.

(a) Sending notice object in Posting Client

(b) Posting 4 Notice Objects

Fig. 5. Implementation results.

4 Conclusions

In this paper, we proposed a noticeboard with EPD for effectively managing notice objects from various users and minimizing power consumption. The system posts more important and newer information based on authority level of the user, posting term and priority of the notice object. Because noticeboard with EPD requires power supply when data is updated, the power consumption is relatively low compared to LCD or OLED based noticeboard systems. We also showed the feasibility of the proposed system by example implementation for managing digitalized notice objects in Korean university. E-paper display is quite suitable to be used for applications requiring low power and mainly using image data. We expect that our study shall be applied to mobile billboards, label, post-it and so on.

References

1. Peng, F.B., Robert, A.H., Ming, L.J., Ling, L.S., Zi, C.Y., Wang, L., Xiao, Z., Guo, F.Z.: Review of paper-like display technologies. Prog. Electromagnet. Res. **147**, 95–116 (2014)
2. Comiskey, B., Albert, J.D., Yoshizawa, H., Jacobson, J.: An electrophoretic ink for all-printed reflective electronic displays. Nature **34**, 253–255 (1998)
3. Nehani, J., Brunelli, D., Magno, M., Sigrist, L., Benini, L.: An energy neutral wearable camera with EPD display. In: Proceedings of the 2015 Workshop on Wearable Systems and Applications, pp. 1–6 (2015)
4. Kim, C.A., Ryu, H.J.: Research trend of the human friendly display-a reflective display. Electron. Telecommun. Trends **28**(5), 1–10 (2013)
5. Lee, S.S., Park, M., Lim, S.H., Kim, J.K.: Electrophoresis-based E-paper display. KIC News **13**(3), 1–13 (2010)
6. LinkSprite. http://www.linksprite.com

Design of Processing Model for Connected Car Data Using Big Data Technology

Lionel Nkenyereye and Jong Wook Jang(✉)

Department of Computer Engineering, Dong-Eui University,
176 Eomgwangro, Busanjin-Gu, Busan 614-714, Korea
lionelnk82@gmail.com, jwjang@deu.ac.kr

Abstract. Recently, we have witnessed a period which things are connected to the Internet. Connected cars are currently among things connected to the Internet. Wireless communications technologies built-in or brought in connected cars enable data generated by in car sensors to be transmitted to external computers where it is analyzed. The main challenge for connected cars services providers is that the collection of same vehicle's data such as engine temperature, engine Revolutions per minute (RPM), vehicle speed are subjected to different connected cars applications which the final purpose of each of them differs. This paper studies design steps to take in consideration when implementing Map Reduce patterns to analyze vehicle's data in order to produce accurate useful outputs. These outputs obtained through big data technology forms a storage repository for the automakers and connect cars services providers. The proposed analytical model is based on a data-driven approach. This approach consists of collecting data sets uploaded from connected cars. Those data are then monitored based on different aspects of activity of the vehicles that we quote as "Events". Hadoop supplements by Map-Reduce functions based reduce side joins with One-To-One joins has been deployed to process a large data and delivered useful outputs. The outputs merged with external information constitute a great insights to connected cars in order to afford connected cars applications.

Keywords: Connected car · Hadoop project · Big data problem · Map reduce · Hadoop · Join algorithms · On-Board diagnostics (OBD-II)

1 Introduction

The information technology authorizes information sharing between vehicle and driver, vehicle sensor operation vehicle surrounding and road condition [1]. Based on this data, the third party interested in automobile ecosystem in occurrence car manufacturers, repair shops, road and transportations authorities will continuously support car user's needs which increase safety driving for the driver but also provide for free or pay as the car owner subscribes to the service. Connected car is a cloud base vehicle information platform [1]. Therefore, the platform is highlighted to cover business operations and services. The wireless communication technologies built-in or brought in the vehicle enable in-car telematics application such as Remote On-line Vehicle Diagnostics System(ROVDS) [2] to send data generated in-car sensors or through OBD connector to external computers.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_23

The information technology expands challenges of ROVDS that submerge currently acquisition of information from what is captured, processed and stored [3]. Such a system might store 40 TB of status and conditions over a year in case connected vehicles reporting Diagnostic Troubles Codes (DTC) status in the size of 25 GB on a daily basis [3].

The collection of vehicle's data allows the third party interested in automobile ecosystem for continuously support new in-car telematics services. Automobile ecosystem is composed by car manufacturers, repair shops, road and transportations authorities. The new in-car telematics services increase safety driving for the driver but also provide for free or pay as the car owner subscribes to these telematics services. Therefore, in order to provide great market for affording these in-car telematics services, the reality of big data technology was explored at a recent Connected Vehicle Trade Association (CVTA) [4].

The Big data technology allows to manage effectively four keys attributes of big data [5]. The first key is the volume that is the quantity of data to be captured which continues to grow exponentially. The second key is velocity that means bits and bytes have to be processed at high speed. The third key is variety which means data comes in many formats, from diverse sources. Finally, value key attribute that emphasizes how data needs to be converted into meaningful insights.

The main challenge for connected cars services providers is that the collection of same vehicle's data such as engine temperature, engine Revolutions per minute(RPM), vehicle speed are subjected to different connected cars applications which the final purpose of each of them differ as shown on the Fig. 1. The automakers and/or connected cars services providers have to design an efficient monitoring and analytics framework. The data processing model is designed using Map Reduce framework which provides a highly analytics of vehicle' data in order to produce accurate useful meaningful insights. These meaningful insights enable connected car providers to target new in cat telematics services related connected car applications. The Fig. 2. shows the system model for implementing architecture for Data Lake using big data technology [5].

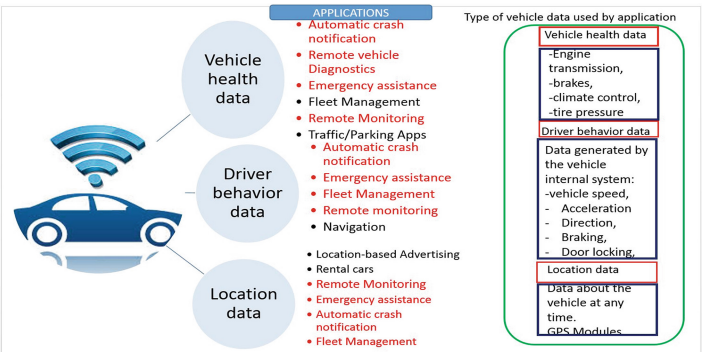


Fig. 1. Vehicle's data that enable Connected Car applications

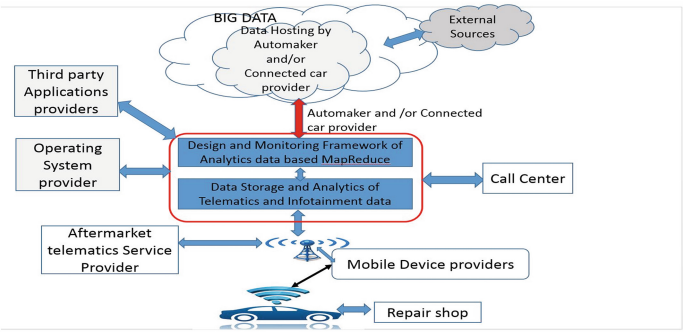


Fig. 2. System model for implementing architecture for Data Lake using big data technology

The main contribution of this work states a processing model based big data technology for monitoring vehicle's diagnostics data and building up a data warehouse accessible to automobile ecosystem. The framework analytical model is expressed as a function that takes input from the Hadoop Data File System (HDFS) on which the Map Reduce join reduce side algorithm is applied to get the final output. The analysis keeps splitting out the vehicle's data based on events occurred inside the vehicle, then select the key value pairs to include in the input functions which are submitted to Map Reduce jobs implemented on the Hadoop cluster.

The remaining of this paper is organized as follows. In Sect. 2, we provide the data processing model for connected car data using big data. In Sect. 2.1, we provide the implementation of the proposed processing model. In Sect. 3, we present conclusions and discuss future research directions.

2 Data Processing Model for Connected Car Data Using Big Data Technology Map Reduce

2.1 Monitoring and Analytical Model Based on Map Reduce

The monitoring and analytics framework is based on a data-driven approach. This approach consists of collecting data sets uploaded from connected cars. These data are then monitored based on different aspect of the vehicle which we denote in this paper as events. Data related to each event is about collection of vehicle's diagnostic data while the driver is driving. The system design of the proposed analytics framework associates for both events vehicle's movement, journey trip, location based service, over speed, mileage and diagnostics of car's engine. At each event, an appropriate subset of information is associated. For instance, for vehicle state and journey trip, the subset of information are journey data, Global Positioning System (GPS) data, driver behavior data based on the smartphone's in- built accelerometer, engine data and car diagnostics data as shown on the Fig. 3. For instance, the analysis process takes the vehicle's movement and journey trip event, then associates in turn RPM value of the vehicle to detect if the engine is running, current data, and accelerometer data to detect

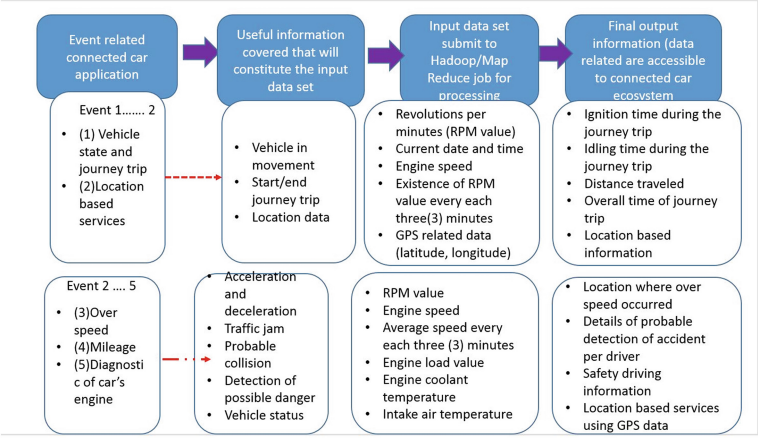


Fig. 3. Design of Monitoring and Analytics framework of Vehicle's data for based on Big Data technology for implementing Remote Vehicle Diagnostics services

vehicle's movement. For this event, the analysis process extracts the value of the RPM every three (3) minutes to detect the state of the vehicle, either is in idle state or not. These data are then uploaded to the Hadoop Data File System (HDFS). The data set on the HDFS serves as the basis for collecting the useful information to submit to Map Reduce functions for processing.

The Hadoop framework splits the input data-set into multiple chunks, each of which is assigned a map task that can process the data in parallel. Each map task reads the input as a set of key-value pairs and produces a transformed set of key-value pairs as the output. The framework shuffles and sorts outputs of the map tasks, sending the intermediate key-value pairs to the reduce tasks, which group them into final results. For the vehicle's movement and journey trip the results are for example vehicle movement time, idling time, traveled time, journey trip time. These results are then stored on the hosting database with an additional field to indicate on which vehicle's telematics applications the output are intended to be applied.

The captured data sets are uploaded using tethered connectivity models that stands on the obligation of carrying the smart phone inside the vehicle. This smart phone is used as a modem via Wi-Fi. When on-board diagnostics data are uploading to the database, Apache Sqoop [6, 7] performs a replication import of data required to run Map Reduce functions. Using HIVE [7] has an important role especially for data stored unto a relational database. Sqoop generates a Hive table based on table originally relational data source and at the same time stores data on HDFS [7].

One of the most key of Big Data technology in this paper is to take the collection of on-board diagnostics and process them according to the final output of useful information and in turn writes back the processing outcomes to the MySQL database. This is achieved by using HIVEQL and Map Reduce functions.

2.2 Distributed Computing with Map Reduce Using Reduce Side Join on Mote Two Data Sets Features One-to-One Join

Map Reduce is a framework for performing analysis of HDFS files in parallel across large dataset using a large number of nodes (computers) [7]. The input data to be computed were so large that require a distributed processing over hundreds or thousands of nodes known as clusters. In this paper, reduce side join on more two data sets features One-to-One joins is used [7].

Figure 4 explains a pseudo code for reduce side join on more two data that features One-To-One join. A list of data sets (tables) on which we want to process is passed as part of the job configuration for allowing the Mapper and Reducer to know how many data sets to expect and what tags are related to each table. The Mapper reads tables, then points tags out from the job configuration and invokes the map function. The join key is removed from list value in order to form a list which includes all the tuples from all the data sets, then rejoining after with the tag of either table T1 or T2 or Tn in a single list value. The tuples based on the dataset they originate from are then proceed by the map function. After Partitioner and the Grouping function have partitioned and grouped the tuples from the all data sets by taking in consideration just the key, and ignoring the tag, the Reduce function gets the tuples sorted on the (key, tag) composite key. All tuples having the same value for the join key are perceived by the same Reducer and only one Reducer function is invoked for one key value. During execution

```
Map (K: intvalue, V from table1 T1 or table2 T2)
  joint_key=return the join column from V;
  listvalue=splitting data and retrieve a list of
value from table1 or table2;
  rm_joint_key =function remove()← remove
the joint_key from listvalue;
 record_listvalue=re-jointing record from
listvalue into a single string by adding a tag of
either table1 or table2;
 map_tag_key=set the joint_key;
 write(map_tag_key, record_listvalue)
```

```
Reduce(K'' : map_tag_key,
LIST_V''(record_listvalue:records from table1
and table2 with map_tag_key K'')
   Create temporary_buffer TB_t1 and TB_t2 for
table1 T1 and table T2 respectively
     for each_record m in LIST_V'' do
        add m to one of the temporary_buffer
TB_t1 and TB_t2 according to tag table
     for each pair of records(t1,t2) in TB_t1 x TB_t2
do
     write_output(null,new_record(pair of
records(t1,t2)))
```

Fig. 4. Pseudo code for Reduce side join on more two data set features One-To-One Join

of Reducer function, the Reducer creates buffers to hold all list values, then groups them within their composite key dynamically. The tuples for a particular key are divided by their parent datasets. A cartesian product of these tuples is performed. Finally, the joined tuples are written to the output.

3 Conclusions and Future Works

With the development of Hadoop platform project, now is possible to build big data solution using open source projects integrated with Hadoop. In this paper, we design a monitoring and analytics framework based on Big Data technology for processing data from vehicles using Hadoop. The final results are written back to a database where connected cars providers can access them via web services.

In Future work, we will focus on the implementation and evaluation of performance of Remote Vehicle Diagnostics service and specific Connected Car application on cloud platforms for instance EC2 (Amazon Elastic compute Cloud).

Acknowledgments. This work was supported by the Brain Busan 21 Project (2016), Nurimaru R&BD project (Busan IT Industry Promotion Agency, in 2016) and Research Institute of Dong-Eui University (2016).

References

1. Whaiduzzaman, M., Sookhak, M., Gani, A., Buyya, R.: A survey on vehicular cloud computing. J. Netw. Comput. Appl. **40**, 325–344 (2014)
2. Kimley-Horn and Associates Inc.: Traffic Management Centers in a connected vehicle environment. Future of TMCs in a connected vehicle, pp. 1–27 (2013)
3. Michigan Department of Transportation and Center for automotive research: Connected Vehicle Technology Industry Delphi Study (2012). http://www.cargroup.org/assets/files/mdot/mdot_industry_delphi.pdf
4. Amanda, M.D.: The art of possibility: connected vehicles and big data analytics (2014). http://blogs.sas.com/content/customeranalytics/2014/12/29/the-art-of-possibility-connected-vehicles-and-big-data-analytics/
5. Cui, B., Mei, H., Chin, B.O.: Big data: the driver for innovation in databases. Nat. Sci. Rev. **1** (1), 27–30 (2014)
6. Jiang, D., Tung, A.K.H., Chen, G.: MAP-JOIN-REDUCE: towards scalable and efficient data analysis on large clusters. IEEE Trans. Knowl. Data Eng. **23**, 1299–1311 (2011)
7. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**, 107–208 (2008)

Efficient Path Selection for IoT Devices in Heterogeneous Service Environments

Dae-Young Kim[1] and Seokhoon Kim[2(✉)]

[1] Department of Software Engineering,
Changshin University, Masanhoewon-gu, Korea
kimdy@cs.ac.kr
[2] Department of Computer Software Engineering,
Soonchunhyang University, Asan, Korea
seokhoon@sch.ac.kr

Abstract. Internet of Things (IoT) services are based on information of IoT devices adopted various sensors. The sensed data by the sensors of devices is delivered to a data server, which is placed in Cloud, for intelligent services. IoT devices serve heterogeneous services using the data. For the connectivity of the IoT devices, wireless sensor networks are exploited as an access network technology for the services. In the wireless sensor networks, energy efficiency is the important issue. Thus, the efficient path selection should be considered. Data transmission in heterogeneous service environments employs aggregation and piggybacking functions. To maximize the energy efficiency, the functions are properly used for data delivery. In this paper, similarity of transmission data in devices is exploited. By the similarity value, the efficient path is selected and one of the functions is used. That is, this paper proposes the path selection to mainly use the aggregation function by high similarity in order to obtain the energy efficiency.

Keywords: IoT · Path selection · Sensor network · Routing · Network service

1 Introduction

Nowadays, Internet of Thins (IoT) technologies lead to information communication technology (ICT) services. IoT is expended to various areas by converging to the area such as industry, transportation, agriculture, military, and so on. For intelligent services, heterogeneous service environments are constructed. IoT devices make a network to collect information in the environments. Various traffics with different characteristics are generated and are delivered over the network. The network of the IoT devices is a wireless sensor network. The wireless sensor network is the access network for IoT services and information is collected through the network [1–4].

The wireless sensor network is composed of various sensors. They measure their neighbor environments and deliver the measured information to a data server. Because IoT devices as the sensors have small battery resources, energy efficiency is the important issue in the wireless sensor network. In addition, lots of data will be transmitted over the network. The data transmission has large portion of the energy

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_24

consumption in IoT devices. Therefore, energy-aware communication should be provided in the wireless sensor networks [5–7]. To minimize energy consumption during data transmission, proper routes should be selected. The routes can have the shortest path or can deliver data with proper size in order to reduce the transmission energy. That is, for the energy efficient transmission, transmission counts are reduced or the amount of transmission data is reduced.

In this paper, the manner to reduce the amount of transmission data is considered. If a path with similarity of transmission data is selected, data can be aggregated in the next hop. Otherwise, data will be piggybacked in the next hop. When the piggybacking function is used, the size of delivered data is increased and more energy is consumed. Thus, the proposed method defines the similarity to select a path. According to the similarity, it chooses the next hop. Then, it can maintain the data size by the aggregation function. Because there are various different data in heterogeneous service environments, if similar data can choose the same path, transmitted data size can reduce. The proposed method aims to do that. Then collecting information in the IoT access network can be optimized.

2 Background

For personal services of IoT, there exists several communication technologies such as WiFi, Bluetooth, and ZigBee. They use unlicensed frequency band and are exploited for various goals. Especially, for IoT networking, the importance of wireless sensor network technologies such as ZigBee is raised [8]. IoT connects numerous devices and gathers their information. And then, it exploits the information for smart services. The network architecture and the service model of IoT is very similar to the wireless sensor network. Thus, the wireless sensor network is known to the access network for IoT.

Figure 1 shows the network architecture of the wireless sensor network for IoT. In the network to collect information, numerous data traffic is generated and is delivered using multi-hop transmission. In addition, the IoT devices as sensor nodes are embedded

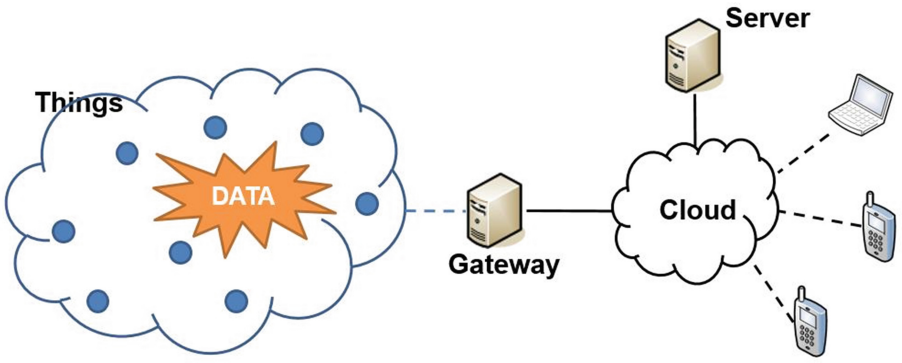


Fig. 1. Network architecture.

systems with limited resources or battery-driven portable devices. Thus, efficient data transmission should be provided in the wireless sensor network in order to maximize the lifetime of the IoT networking.

For the efficient data transmission, several schemes are considered. First, by finding the shortest path for data transmission, energy consumption can be reduced. Second, by reducing data size to transmit, the energy consumption can be also reduced. As mentioned earlier, in this paper, we are interested in the data size to reduce energy consumption. That is, the proposed method in this paper selects the next hop which is possible to aggregation instead of piggybacking. By the aggregation, the proposed method can reduce the transmission energy.

3 The Proposed Path Selection Scheme

3.1 Radio Model

In the wireless sensor network, there is a radio model to represent the transmission energy consumption at sensor nodes. The radio model calculates the energy consumption of a sender and a receiver using transmitted data size and a distance between a sender and a receiver [9, 10]. It can be described as the following equation.

$$\begin{aligned} E_{TX}(n,l) &= E_{elec} \times n + \varepsilon_{fs} \times n \times l^2 \\ E_{RX}(n) &= E_{elec} \times n \end{aligned} . \tag{1}$$

E_{TX} is the consumed energy at a sender when the node transmits n-bit data in a distance l. E_{RX} is the consumed energy at a receiver when the node receives n-bit data. E_{elec} is the consumed energy in electric circuits when data transmission occurs. ε_{fs} is the consumed energy in the amplifier. Because data transmission in the wireless sensor network happens in short transmission range, ε_{fs} is considered as free space model. By the radio model, considering data size and transmission distance is important for the energy-aware data transmission.

3.2 Path Selection

As mentioned earlier, data can be aggregated or piggybacked at intermediate nodes. The data of the same type is aggregated and the data of different types is piggybacked. Because piggybacked data increases data size to transmit, the data aggregation should be mainly performed for efficient data delivery.

To find the path that is possible to aggregate data, the proposed method defines similarity. The similarity is represented by Euclidian distance of generated data at a node. It can be described by kNN algorithm. Each node calculates the similarity for its neighbor nodes. Thus, when a node has data for transmission, it compares the similarity and selects the next hop with the shortest Euclidian distance of data for the data aggregation.

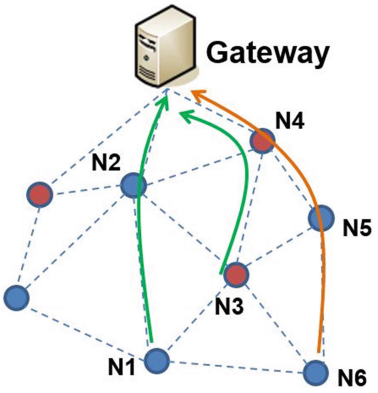


Fig. 2. The proposed path selection scheme.

Figure 2 shows the proposed path selection scheme. In the figure, read and blue nodes generate different data. Green paths are the paths for the same data. In the green paths, data aggregation occurs and data size for transmission is maintained. However, in the yellow path, two types of data are generated. In N6-N5, data aggregation occurs but in N5-N4 data piggybacking occurs and data size for transmission is increased. By the radio model in Sect. 3.1, it causes more energy consumption for data transmission.

```
Path-Selection:
1. Calculate Euclidian distances for generated data of neighbors
2. Classify the shortest path group
3. Loop in the shortest path group
4.     Compare the Euclidian distances (similarity)
5.     Find the shortest Euclidian distance
6.     Select the next hop
7. End loop
8. Transmit data to the selected node
```

Fig. 3. The pseudocode for the proposed scheme.

Figure 3 represents the pseudocode for the proposed scheme. Each node operates the algorithm to transmit data. In wireless sensor networks, a sink node broadcasts query messages to construct routing path. Thus, the proposed scheme finds a next hop with high similarity among the shortest path for the sink node. In the node with high similarity, possibility of the data aggregation is high and thus transmission energy may not be increased. By the proposed method, data piggybacking is reduced and energy-aware data path can be provided to IoT sensor devices.

4 Conclusion

In IoT networking, reducing transmission energy consumption is an important problem because most of IoT devices are tiny embedded systems and they employ battery for their operation. Data size for transmission is major factor in aspect of transmission energy consumption. For IoT services, various data is generated in the IoT area. A server gathers data in the area and the data can be delivered through multi-hop transmission. In the intermediate nodes, the data can be aggregated or piggybacked. To maintain the data size for transmission, data aggregation should be performed instated of piggybacking. The proposed method deals with the path selection to increase the possibility of the data aggregation.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2060035) and this work was supported by the Soonchunhyang University Research Fund (No. 20160220).

References

1. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Elsevier Comput. Netw. **54**(15), 2787–2805 (2010)
2. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): A vision, architectural elements, and future directions. Elsevier Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
3. Kim, S., Na, W.: Safe data transmission architecture based on cloud for Internet of Things. Wireless Pers. Commun. **86**(1), 287–300 (2016)
4. Bandyopadhyay, D., Sen, J.: Internet of Things: applications and challenges in technology and standardization. Wireless Pers. Commun. **55**(1), 49–69 (2011)
5. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirei, E.: A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
6. Culler, D., Estrin, D., Srivastava, M.: Guest editors introduction: overview of sensor networks. IEEE Comput. **37**(8), 41–49 (2004)
7. Kim, D.-Y., Jin, Z., Choi, J., Lee, B., Cho, J.: Transmission power control with the guaranteed communication reliability in WSN. Int. J. Distrib. Sensor Netw. **2015**(5), 1–12 (2015). ID632590
8. ZigBee Alliance. http://www.zigbee.org/
9. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy-efficient communication protocol for wireless sensor networks. In: Proceedings of the Hawaii International Conference System Sciences, January 2000
10. Lindsey, S., Raghavendra, C.S.: PEGASIS: Power Efficient GAthering in Sensor Information Systems. In: Proceedings of IEEE Aerospace Conference, March 2002

Forecasting Sugarcane Yield Using (μ+λ) Adaptive Evolution Strategies

Supawadee Srikamdee, Sunisa Rimcharoen, and Nutthanon Leelathakul(✉)

Faculty of Informatics, Burapha University, Chonburi, Thailand
{srikamdee, rsunisa, nutthanon}@buu.ac.th

Abstract. Sugarcane is a very important crop in the sugar industry. However, the annual amount of harvested sugarcane is oftentimes uncertain, posing risks to sugarcane mills in terms of raw material supply. The forecast for the sugarcane yield would allow the mills to plan sugar production accordingly. This paper proposes (μ+λ) adaptive evolution strategies, which generate equations for accurately forecasting the sugarcane yield. Our proposed scheme combines the advantages of two algorithms: genetic algorithm and evolution strategies. Specifically, the genetic algorithm is good for determining patterns of forecasting equations, while the evolution strategies are used to tune the equations' coefficients. The test data is collected from sugarcane farmers in 24 provinces of Thailand during 2010–2014. The equations obtained by the proposed method are 80 % accurate on average, outperforming the previous method (back propagation neural network) in all data set.

Keywords: Evolution strategies · Forecast · Genetic algorithm · Sugarcane

1 Introduction

Sugarcane is one of Thailand's important industrial crops. It is a raw material for sugar production, and can also be used as an alternative energy source. However, sugarcane farmers could not guarantee their yield to the mills, which have to take risks of lacking raw material supplies. Therefore, the ability to forecast the sugarcane yield, accurately and consistently with reality, is crucial to the mills' production planning for domestic consumption and/or export.

The sugarcane yield each year is uncertain due to various factors, e.g. species, cultivation methods, caretakers, and harvest methods. There are two ways to grow sugarcane: from new sprouts (setts) and from stubbles. When newly planted sugarcanes are one year of age, they are ready to be harvested. Farmers would cut the cane stems, and leave the roots and the lower parts (i.e. stubbles) uncut. Then, the farmers plough the soil and add fertilizers in order to regrow canes from the stubbles. The stubbles can be used to regrow canes no longer than three times because the canes get thinner and thinner after each harvest, and develop diseases more easily. We call the canes after subsequent harvests first ratoon, second ratoon, and so on. In general, Thai farmers plant new sprouts (new setts) after harvesting the second or third ratoons.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_25

Forecasting sugarcane yield is not straightforward due to the uncertainty of various factors, e.g. weather, irrigation, pests, a plantation area, and cultivation. Researchers in [1] collected the information related to sugarcane farming in the Northeast of Thailand during 2004–2009. The study found factors correlated with the sugarcane yield, which are province, year, average rainfall, yield per rai (i.e., 1600 square meters), and a plantation area. The model has correlation coefficients of 0.99.

In general, sugarcane mills calculate tentative yield by multiplying a plantation area (obtained from the quota database) with average yield per area. This method is simple but quite inaccurate. Thus, other research works attempted to create forecasting models that are more accurate. For instance, the researchers in [2] applied popular neural networks to forecast the yield of each plantation area. A mathematical model (with MAPE of 19.5 %) was proposed to recommend the optimal harvest schedule. To forecast sugarcane yield in Nigeria, the study in [3] applied the neural networks to the data, taking into account economic and environmental factors during 1920–2005. The prediction accuracy is 85.70 %.

Although the neural networks in [2, 3] are successful in forecasting the sugarcane yield, their black box nature makes it hard to understand and interpret. (Neural network models normally return the set of neuron-node coefficients, not equations, which general individuals could comprehend.) In this paper, we propose (μ+λ) Adaptive Evolution Strategies for speculating the sugarcane yield, which will be supplied to the mills. The equations, generated by our proposed method, consist of related factors together with their coefficient.

2 The Adaptive Evolutionary Strategies

The adaptive ES (A-ES) is an algorithm that incorporates the advantages of two evolutionary algorithms: genetic algorithm (GA) and evolution strategies (ES). It could be used to determine forecasting equations together with coefficients. Particularly, GA's capability of evolving functional forms is combined with ES' capability of tuning the coefficient values. The A-ES algorithm succeeds in generating equations for forecasting (1) exchange rates for Thai Baht to US Dollar in [4], and (2) Thai SET index in [5].

ES is often used to search for equations' coefficients, or to finely tune the parameter values when the forms of the equations are already known. However, at times, the equation forms might still be in question. For example, we might want to find an equation form that fits a given set of data. Genetic Algorithm (GA) or Genetic Programming (GP) is one of the popular methods for performing such a task. Both choose coefficient values at random; thus, yielding unreliable values in some cases [6]. In contrast, ES provides more accurate coefficient values since it gradually improves the coefficient values instead of changing them radically.

This paper proposes (μ+λ) A-ES to generate equations for forecasting sugarcane yield, being supplied to the mills. Variables $term_1$, $term_2$,…, $term_n$ in the forecasting equations represent input data related to each farmer's sugarcane planting, e.g. a plantation area, sugarcane species, and rainfall. The outputs of the equations are predicted values of sugarcane yield. In this work, we improve the previous work [5] by

adding the step of crossover (i.e., the step 2.3 in the procedure below). The chromosome encoding (shown in Fig. 1) is also different from the one in [5] in that the number of coefficients is smaller, and the primitive function is simpler.

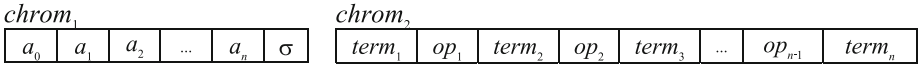


Fig. 1. Chromosome encoding

In Fig. 1, a_0 is a constant; n is a number of terms in the equation; $a_1 \ldots a_n$ are coefficients of each variable; $term_1 \ldots term_n$ are variables; $op_1 \ldots op_{n\text{-}1}$ are operators; σ is a variance used in controlling the search strategy. In this work, the procedure of the (μ+λ) A-ES is similar to the one in [5], except the step 2.3, which is shown as follows.

Step 2.3 Crossover the functional form ($chrom_2$) by randomly select a crossover position and exchange the chromosome structures. Mutate the functional form ($chrom_2$) by changing a function or an operator to another function or operator at a random position.

$$chrom_2' = \text{crossover}(chrom_2); \qquad chrom_2'' = \text{mutate}(chrom_2')$$

3 Experiments

3.1 Input Data

The input data collected for this work are the information obtained from farmers in 24 provinces of Thailand, which supplied 8 – 14 tons of sugarcane to the mills during 2010–2014. The data points, whose yield is more different (i.e., more than 30 % different) from the quotas the farmers agree with the mills, are excluded. The numbers of the farmers, contributing to the data during 2010–2014, are 314, 284, 288, 312 and 220, respectively.

The data consist of 9 attributes: areas of new sprout plantation (rais), areas of first-ratoon plantation (rais), areas of second-ratoon plantation (rais), areas of third-ratoon plantation (rais), quota agreements with the mills (tons), species (i.e. K84-200, LK11, K88-50, K90-77, K88-92, K88-50, OoTong3, and OoTong1), average rainfall of plantation (millimeters), number of rainy days throughout the year of each plantation (days), and maximum rainfall of each plantation (millimeters). Our proposed A-ES runs on the data of the 9 attributes to create a forecasting model to speculate the total sugarcane yield (tons) of both with and without pre-harvest burning.

3.2 Experimental Design

We use Mean Absolute Percent Error (MAPE) to measure the performance of our proposed method.

In this experiment, the parameters of A-ES are set as follows. The number of parents and children (μ and λ) are 50 and 100, respectively. The form mutation rate and the coefficient mutation rate are 0.1. The crossover rate is 0.6–0.8.

We compare the performance of our method with the one of Back-Propagation Neural Networks (BPNN). We use the BPNN implementation in Weka 3.6, and set its parameters as follows. The number of hidden nodes is 16, which performs best in the experiments among the range of 5–20 hidden nodes. The learning rate is 0.1. The momentum is 0.1 and the number of iterations is 3000, which performs best in the experiments among the values of {1500, 2000, 2500, 3000}.

3.3 Experimental Results

This subsection presents the performance comparison between our proposed method and BPNN. We conduct two experiments. The former is for assessing short-term forecast where the data of year *t* are used to create the model for forecasting the yield of year $t + 1$. The latter is for evaluating long-term forecast where only the data of year 2010 are used to create the forecasting model for the yield of each year during 2011–2014.

3.3.1 Short-Term Sugarcane Yield Forecast

The MAPE of A-ES and BPNN, running on training and testing data, are shown in Table 1. In the training process, the MAPEs of our method (A-ES) are in the range of 9.64 %–11.65 %, while the ones of BPPN are in the range of 10.27 %–84.31 %. Likewise, the comparison results are similar in the testing process. On average, the models obtained from A-ES are better at short-term forecasting for the sugarcane yields by 43 %.

The equations in the last column of Table 1 suggest the major factors that could have an impact on the forecast accuracy. We can see that the factors – (1) plantation areas of new-sprout crops, (2) ones of first-ratoon crops, (3) ones of second-ratoon crops, (4) ones of third-ratoon crops, and (5) quota agreements with the mills - are taken into account in almost all equations. The fact is confirmed once we analyze their correlation values, which are 0.60, 0.62, 0.72, 0.66, and 0.99, respectively. The factors we found are also consistent with the simple equation (i.e. yield = area × yield per area), which the mills use to compute tentative yield [7]. However, the yield per area are rarely accurate as it depends on various factors (e.g., weather, species, and cultivation). The equations derived by our proposed method (A-ES) take into account more factors and more specific in terms of the stubbles' year of age, thus yielding more accurate predictions.

3.3.2 Long-Term Sugarcane Yield Forecast

In Table 1, the first equation, where the training data is of year 2010, yields the lowest MAPE. As a result, we choose this equation to be our model for forecasting the yield in the long term. Specifically, we use the data during 2011–2014 to be the testing data sets in order to measure the accuracy of this equation. The MAPE results of both A-ES and BPNN are shown in Table 2.

During the training period using the data of year 2010, the MAPE of A-ES (9.64 %) and the one of BPNN (10.27 %) are quite close. However, when we use the two methods to predict the sugarcane yield during 2011–2014, the MAPEs of the two are farther and farther apart, as shown in Table 2. The MAPEs of A-ES continue to be low and are similar to their average (12 %, approximately) while the ones of BPNN keep increasing. The results show that A-ES can generate the accurate models for both short-term and long-term forecasts. The A-ES model once generated could forecast the sugarcane yield accurately at least for the next 4–5 years. Therefore, it is unnecessary to train the forecasting model so often.

Table 1. MAPEs of the proposed model (A-ES) and BPNN (Short-term forecast)

Train	Test	MAPE (train)		MAPE (test)		Equation
		A-ES	BPNN	A-ES	BPNN	
2010	2011	9.64	10.27	12.43	20.08	$0.9 + 3.4Ratoon3 + 3.61Ratoon1 - 8.8Type_{K88-50} - 0.15Type_{OoTong1} * 1.4Quota + 3.23Planted + 0.7Quota - 7.73Type_{K84-200} + 3.45Ratoon2$
2011	2012	11.65	18.56	18.08	45.98	$2.4Ratoon1 + 1.83Type_{LK11} + 3.1Ratoon3 + 2.39Ratoon2 * 3.25Type_{OoTong1} * 7.9Type_{K88-50} + 2.5Type_{K88-92} + 0.72Quota + 2.79Planted$
2012	2013	10.93	84.31	11.03	77.62	$7.82 + 0.39Quota + 5.67Ratoon2 + 5.55Ratoon3 + 5.85Planted - 8.44Type_{LK11} * 1.04Ratoon3 + 5.39Ratoon1$
2013	2014	11.42	38.45	11.29	81.73	$2.69 + 9.45Ratoon3 + 10.16Planted - 3.09Type_{K88-50} + 9.54Type_{K84-200} * 0.01Ratoon2 + 60.29Type_{K88-92} + 9.44Ratoon1 + 9.34Ratoon2$
Average		**10.91**	**37.90**	**13.21**	**56.35**	

Table 2. MAPEs of the proposed model (A-ES) and BPNN (Long-term forecast)

Year	2011	2012	2013	2014	**Average**
A-ES	12.43	12.81	12.37	12.41	**12.50**
BPNN	20.08	35.95	37.51	52.32	**36.47**

4 Conclusion

This paper presents an algorithm for creating a model to forecast sugarcane yield, which Thai farmers would supply to the mills. We collected the data from sugarcane farmers in 24 provinces of Thailand during 2010–2014. Our forecasting equations

include the following factors: sugarcane cultivation areas, species, the quota each farmer agreed with the mills, and rainfall. The experimental results show that our proposed algorithm outperforms BPNN in all study cases, in both short-term and long-term forecasts. From the obtained equations, major factors, contributing to the sugarcane yield, are (1) the cultivation areas of the new-sprout, first-ratoon, second-ratoon, and third-ratoon crops, and (2) the agreement quotas. To improve the accuracy, more factors, e.g. cultivation duration, farmer skills, temperature, and moisture, should be taken into account.

Acknowledgments. This work was financially supported by the Research Grant of Burapha University through National Research Council of Thailand (Grant no. 52/2559).

References

1. Bugate, O., Seresangtakul, P.: Sugarcane production forecasting model of the northeastern by artificial neural network. KKU Sci. J. **41**(1), 213–225 (2013)
2. Thuankaewsing, S., Pathumnakul, S., Piewthongngam, K.: Using an artificial neural network and a mathematical model for sugarcane harvesting scheduling. In: The IEEE International Conference on Industrial Engineering and Engineering Management, pp. 308–312 (2011)
3. Obe, O.O., Shangodoyin, D.K.: Artificial neural network based model for forecasting sugar cane production. J. Comput. Sci. **6**(4), 439–445 (2010)
4. Rimcharoen, S., Sutivong, D., Chongstitvatana, P.: Curve fitting using adaptive evolution strategies for forecasting the exchange rate. In: The 2nd ECTI Annual International Conference (2005)
5. Rimcharoen, S., Sutivong, D., Chongstitvatana, P.: Prediction of the stock exchange of Thailand using adaptive evolution strategies. In: The 17th IEEE International Conference on Tools with Artificial Intelligence, pp. 232–236 (2005)
6. Koza, J.R.: Genetic Programming: on the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge (1992)
7. Thuankaewsing, S.: Applying Artificial Neural Network and Geographic Information System to Forecast Sugarcane Yields. Master Thesis, Khon Kaen University (2009)

Resource Pooling Mechanism for Mobile Cloud Computing Service

Seok-Hyeon Han, Hyun-Woo Kim, and Young-Sik Jeong(✉)

Department of Multimedia Engineering, Dongguk University, Seoul, Korea
{shhan, hwkim, ysjeong}@dongguk.edu

Abstract. Cloud computing services enable virtual services with physical resources consisting of workstations and clusters having storages, processes, and memories. Unlike general resources, mobile devices have such features as miniaturized hardware, embedded sensors, and touch screen. Furthermore, mobile devices have mobility and portability which are lacking in general workstations and clusters. Mobile cloud computing services organize physical resources with mobile devices which have low hardware performance and are sporadically connected to the network. Therefore, it is difficult to integrate resources if they are organized as in general cloud computing services. Resource pooling management is required to manage and provide the resources of mobile devices dynamically in an integrated manner and provide them to users without work loss. This research proposes Mobile Resource Integration – Manager (MRI-M) which checks and manages the resource status of mobile devices in mobile cloud computing services and considers the organization of resource pooling.

Keywords: Mobile cloud computing · Resource pooling · Cloud computing service · Mechanism · Mobile · Resource pool

1 Introduction

General cloud computing services organize a resource pool through workstations and clusters having storages, processes, and memories, and the services are provided through virtualization using the resource pool. Once a resource pool is organized, they can provide encoding and image processing services which cannot be processed in a single mobile device. The mobile devices of mobile cloud computing services are sporadically connected to the mobile cloud through a wireless network and their hardware specifications are lower than those of workstations and clusters. Therefore, a resource pool management method is required to dynamically integrate, manage, and provide the resources of mobile devices to users with no work loss [1, 2, 6, 7].

In this paper, proposes Mobile Resource Integration – Manager(MRI-M) which checks and manages the resource status of mobile devices in mobile cloud computing services and considers the organization of resource pooling.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_26

2 Related Works

In conventional studies, various services using mobile devices enable mobile cloud computing services, various applications, and the collection of various data [1, 3]. Cloud services have been provided through extended wireless sensor networks from the aspect of applications. This technology can be applied to many fields such as medicine, military, environment, and manufacturing [4]. Furthermore, cloud data centers have been created to control the uses of resources from the center through servers [5–7]. However, this method is inappropriate for mobile cloud computing services that only use mobile devices with no fixed servers. This paper proposes a new resource pooling method that enables mobile cloud computing services using mobile devices.

3 Scheme of MRI-M

The Mobile Resource Integration – Manager (MRI-M) to provide a resource pool for mobile cloud computing services is organized as shown in Fig. 1.

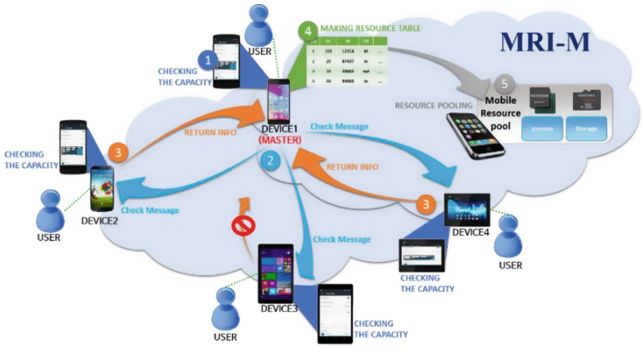


Fig. 1. Overview of MRI-M

The MRI-M that organizes a resource pool required for the mobile cloud computing environment has the following features:

① The capacities of mobile devices are measured using a benchmark program before starting the mobile cloud computing services.
② The mobile devices connected to the mobile cloud are checked to see if they work. The master device sends a check message requesting their state and capacity to all mobile devices connected to the mobile cloud.
③ Upon receiving the check message, mobile devices send their capacity and state information.
④ The master device creates a resource table required to organize a resource pool using the received information.
⑤ A resource pool required for mobile cloud computing is organized with the mobile devices connected to the mobile cloud.

The master device setting for MRI-M that checks the connection and information of mobile devices connected to the mobile cloud follows the flow chart in Fig. 2. A mobile device reads the check message when starting a mobile cloud. If there is a check message, it means that the master device exists. Therefore, the mobile device finishes the process without setting it as master device. If there is no check message, the mobile device selects itself as master device. The master device sends check message to all mobile devices connected to the mobile cloud and receives their information. Then it organizes a resource table using the received metadata.

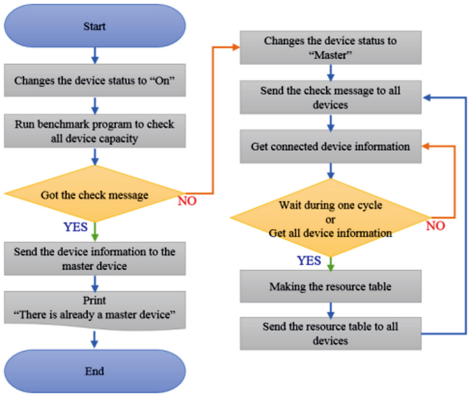


Fig. 2. Master Device Flow chart of MRI-M

The components of the resource table are listed in Table 1.

Table 1. Component of resource table

Component	Explanation
Identity Number (IN)	Indicates the unique identity number of a mobile device
Device Name (DN)	Indicates the name of each device connected to the mobile cloud
Mobile OS(MO)	Indicates the OS of the mobile device
Device Capacity (DC)	Indicates the capacity of a mobile device and is divided into storage and process
Device State(DS)	Indicates the state of a mobile device connected to the mobile cloud (On/Off/Master)
Device Usage (DU)	Indicates the usage capacity of a device through mobile cloud computing

The master device operates in the following sequence to provide a resource pool using a resource table.

① Define the capacity of a mobile device by the storage and process performance. α denotes the storage ratio of the mobile device, and β denotes the process ratio of the mobile device. The sum of α and β does not exceed 100.

$$DC = Device_{storage} \times \alpha + Device_{process} \times \beta$$

② Determine the maximum available capacity and the currently used capacity of the mobile cloud based on the resource table, and calculate the currently available resource capacity through them.

$$_{Total}DC = \sum_{i=1}^{N} DC_i, \; _{Total}DU = \sum_{i=1}^{N} DU_i$$

$$Currently\ Available\ Resource\ Capacity = {}_{Total}DC - {}_{Total}DU$$

$$(where\ N = Number\ of\ Device)$$

③ Calculate the basic resource pool size using the currently available resource capacity, the number of mobile devices, and the device with the smallest capacity.

$$Resource\ Pool\ Size = \frac{Currently\ Available\ Resource\ Capacity - {}_{Min}DC}{N}$$

④ Provide a resource pool to mobile devices in the descending order of mobile devices based on the capacity size. In case the required capacity is smaller than the resource pool size, the job is processed by the mobile device.

4 Design of MRI-M

The functions of MRI-M consist of User Interface, Cloud Pool Manager, Device Manager, Resource Table Manager, Master Manager, Coordinate Converter, and Viewer. Figure 3 shows the architecture of MRI-M.

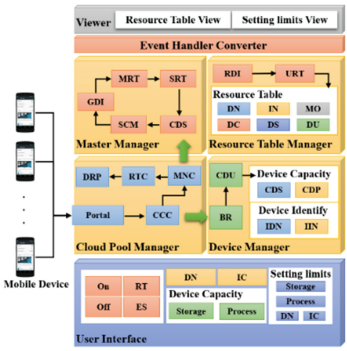


Fig. 3. MRI-M Architecture

User Interface connects users with a mobile device. It consists of Resource Table (RT), Environment Setting (ES), Device Name (DN), IMEI Check (IC), Device Capacity, and Setting limits. Furthermore, Device Capacity and Setting limits are divided into Storage and Process. Device Manager measures the capacity of the mobile device. It consists of Benchmark Run (BR), Check Device Usage (CDU), Device Capacity, and Device Identity. Furthermore, Device Capacity is divided into Check Device Storage (CDS) and Check Device Process (CDP), and Device Identity is divided into Identify Device Name (IDN) and Identify IMEI Number (IIN). Cloud Pool Manager checks the connected devices and provides a pool. It consists of Portal, Cloud Connection Check (CCC), Master Device Check (MNC), Resource Table Calculation (RTC), and Device Resource Pooling (DRP). Master Manager performs operations when the mobile device is the master. It consists of Change Device Status (CDS), Send Check Message (SCM), Get Device Information (GDI), Making Resource Table (MRT), and Send Resource Table (SRT). Resource Table Manager shows the state of mobile devices connected to the mobile cloud. It consists of Receive Device Information (RDI), Update Resource Table (URT), and Resource Table.

5 Implementation of MRI-M

The running screen of the proposed MRI-M is illustrated in Fig. 4. Figure 4-① shows the initial screen of the MRI-M which is used to enter the mobile device name and the IMEI number, and to check the device capacity. Once you check the capacity, press the On button to start cloud resource pooling. Figure 4-② appears when you press the Environment Setting button on the initial screen. You can change the capacity of the mobile device provided when connecting to the mobile cloud. Figure 4-③ appears when you press the Resource Table button on the initial screen. The resource table shows the information of mobile devices connected to the mobile cloud.

Fig. 4. Screen of the MRI-M Implementation

6 Conclusions

In this paper, Mobile Resource Integration – Manager (MRI-M) which organizes and manages a resource pool based on mobile devices to provide mobile cloud computing services. MRI-M checks the addition and removal of mobile devices to/from the mobile

cloud and measures the capacity of each mobile device. Furthermore, it creates a resource table and manages a resource pool using the measured metadata of the mobile devices. Therefore, MRI-M is proposed a resource management method for the organization of the total mobile cloud computing services.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2014R1A1A2053564). And also This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP(Institute for Information & communications Technology Promotion), And also This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No. R7120-16-1006, Resource Integration Management Solution with On-premises Legacy for Intra-Cloud). And also This research was supported by the MISP (Ministry of Science, ICT & Future Planning), Korea, under the National Program for Excellence in SW(R7116-16-1014) supervised by the IITP(Institute for Information & communications Technology Promotion)(R7116-16-1014).

References

1. Aniobi, D.E., Alu, E.S.: Usability testing and evaluation of a cloud computing-based mobile learning app: students' perspective. Int. J. Comput. Sci. Issues **13**(1), 67–75 (2016)
2. Yasmine-Derdour, Y.D., Bouabdellah-Kechar, B.K., Mohammed, F.-K.: Using mobile data collectors to enhance energy efficiency and reliability in delay tolerant wireless sensor networks. J. Inf. Process. Syst. **12**(2), 275–294 (2016)
3. Vehbi, R., Ylber, J., Ymer, J.: Transaction management in mobile database. Intl. J. Comput. Sci. Issues **13**(3), 39–44 (2016)
4. Alamri, A., Ansari, W.S., Hassan, M.M., Hossain, M.S., Alelaiwi, A., Hossain, M.A.: A survey on sensor-cloud: architecture, applications, and approaches. Intl. J. Distrib. Sensor Netw. **2013**(2013) 1–18 (2012)
5. Ong, I., Limm H.: Dynamic load balancing and network adaptive virtual storage service for mobile appliances. J. Inf. Process. Syst. **7**(1), 53–62 (2011)
6. Viswanathan, H., Lee, E.K., Rodero, I., Pompili, D.: Uncertainty-aware autonomic resource provisioning for mobile cloud computing. IEEE Trans. Parallel Distrib. Syst. **26**(8), 2363–2372 (2015)
7. Cui, Y., Lai, Z., Dai, N.: A first look at mobile cloud storage services: architecture, experimentation and challenge. IEEE Netw. **30**(4), 16–21 (2016)

IPC Multi-label Classification Applying the Characteristics of Patent Documents

Sora Lim(✉) and YongJin Kwon

Korea Aerospace University, Goyang, Korea
{ebbunsora, yjkwon}@kau.ac.kr

Abstract. Most of research on the IPC automatic classification system has focused on applying various existing machine learning methods to the patent documents rather than considering the characteristics of the data or the structure of the patent documents. This paper, therefore, proposes using two structural fields, a technical field and a background field which are selected by applying the characteristics of patent documents and the role of the structural fields. A multi-label classification model is also constructed to reflect that a patent document could have multiple IPCs and to classify patent documents at an IPC subclass level comprised of 630 categories. The effects of the structural fields of the patent documents are examined using 564,793 registered patents in Korea. An 87.2 % precision rate is obtained when using the two fields mainly. From this sequence, it is verified that the technical field and background field play an important role in improving the precision of IPC multi-label classification at the IPC subclass level.

Keywords: Patent classification · IPC classification · Patent document fields · Data characteristics · Multi-label classification

1 Introduction

In knowledge-based societies, intellectual property that results from knowledge-related activities has become more important. Globally, vast numbers of patents, each one representative of intellectual property, have already been filed, and the current trend is for continued upward growth. In Korea, there were approximately 210,000 patents filed just in 2014 [1]. Accordingly, it is important to classify patents according to their technical and industrial fields so that patent investigations and patent-related work can be performed effectively. The International Patent Classification (IPC) is widely used for this purpose.

To replace current human-based patent classification processes, research on developing automatic patent classification systems has been conducted using data mining and machine learning techniques [2, 3]. Previous research on IPC-based patent classification has tended to focus on the selection of a machine learning model [4–7] or discriminative words in patent documents [8] rather than to consider data characteristics and their potential benefits.

However, when patent documents are classified, the application of machine learning algorithms alone will not lead to satisfactory results. This is because patent

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_27

documents have their own characteristics that differentiate them from web documents, SNSes and scientific data. Therefore, consideration must be given to particular patent document characteristics as well as to the typical text classification methods when classifying patent documents via an automatic system.

This paper aims to construct an IPC multi-label classification system based on patent document characteristics or the role of the structural fields of patent documents to replace the IPC classification process carried out by humans and to assist the examiner's decision by recommending relevant IPCs at the IPC subclass level (which has 630 categories). This paper also investigates how the structural fields in the patent documents affect the result of the auto-classification.

2 IPC and Patent Understanding

2.1 IPC Structure

The IPC system assigns its own classification codes to patent documents. The inten-tion here is to create an internationally uniform classification system that clas-sifies patents according to their technological areas [9]. The IPC has a hierarchical structure of five levels consisting of eight sections, 128 classes, about 650 subclasses, about 6,800 main groups and more than 65,000 subgroups. See Fig. 1 for the details.

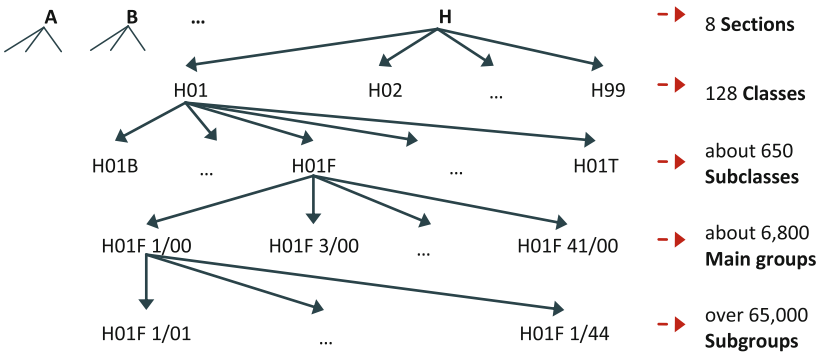


Fig. 1. IPC structure

2.2 Patent Document Characteristics

Patent documents are structured documents that consist of several fields. Figure 2 shows the details. Specially, patent documents can simultaneously possess multiple IPC codes. Given that classification problems are widely divided using binary, multi-class and multi-label classification problems in statistics and machine learning, IPC classification is a multi-label classification problem. For instance, the patent 'Portable terminal device comprising bended display and method for controlling thereof' has two IPC codes – G06F 3/048 and H04B 1/40. The patent 'Manufacturing apparatus and method of solid fuel using food waste' has 11 IPCs – C10L 5/46, A47J

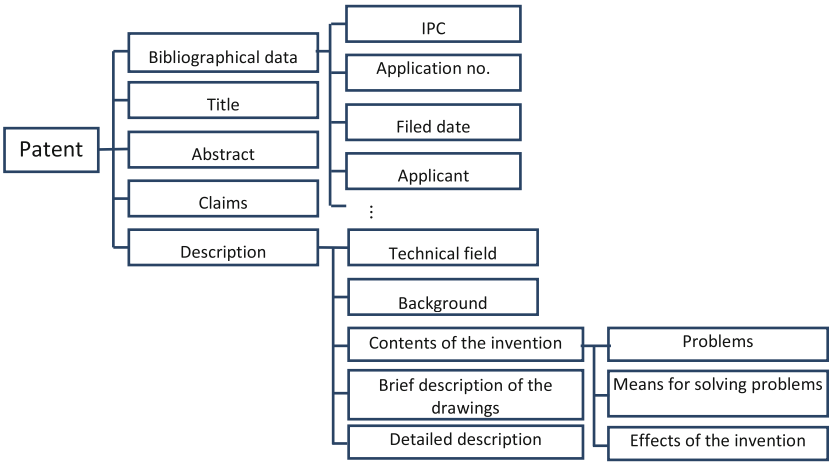


Fig. 2. Patent structure

37/12, B01D 47/00, B01F 7/02, B02C 23/08, B04B 5/10, B07B 1/28, B09B 3/00, B30B 9/04, C10L 5/40 and F26B 3/04.

3 IPC Multi-label Classification

In this paper, patent classification at the IPC subclass level is performed via patent documents collecting, preprocessing, feature selection, vector space modeling, classifier construction and IPC prediction.

Patent Data Collecting. Open API of Kipris Plus [10] is used to collect patent documents registered in Korea from Jan 1, 2010 to Dec 31, 2015. The collected documents are then parsed so that the representative fields, such as titles, abstracts, claims, technical fields and backgrounds are extracted from each of the collected documents to examine their effects on the classification performance. Finally, a patents collection is constructed with 564,793 patents categorized into 630 IPC subclasses.

Preprocessing. Nouns are extracted as the minimum unit with meaning from each patent document in the collection. In this task, a Korean language morphological analyzer, KLT2000 [11] is used. To remove stopwords effectively, including 'invention', 'claim', 'application', 'patent', 'description' and 'drawing', a stopwordlist is constructed based on the index terms in the glossary of the intellectual property terms and patent examination guidelines. There are 1,860 stopwords.

Feature Selection and Vector Space Modeling. Deterministic words (features) are selected for the 630 IPC subclasses by using a TF-ICF scheme [7], which is a variation of the well-known TF-IDF weighting scheme for multi-classifications. Word weights are calculated for each IPC subclass using TF-ICF, as follows

$$TF_{ij} = avg(freq_{ij}),\ ICF = \log_2(N/n_i),$$
$$w_{ij} = TF_{ij} * ICF_i.$$

where TF_{ij} is the average frequency of feature f_i in IPC subclass c_j, and ICF_i is the inverse category frequency of f_i. N is the total number of IPC subclasses used to label patent documents across the entire patent document collection and n_i is the number of IPC subclasses including f_i in the collection. The weight w_{ij} of feature f_i in IPC subclass c_j is subsequently calculated by multiplying TF_{ij} by ICF_i. The top k features are then selected for each IPC subclass. In this paper, it is experimentally determined that when k = 100, the features of each subclass aptly represent the technical subject of the IPC subclass and properly differentiate the subclass from other subclasses. Therefore, a feature set consisting of 33,047 words is constructed after removing overlap. Patent documents are transformed into the feature vectors based on the feature set. The vectors with a row frequency of features are weighted using the TF-IDF [12] scheme.

Classifier Construction and IPC Prediction. To classify patent documents that could have multi-labels, the multinomial naive Bayes model [14] is used. This is one of the most representative machine learning models.

$$P(c|f_1, f_2, \ldots, f_n) \propto P(c) \prod_{i=1}^{n} P(f_i|c)$$

$$\hat{c} = argmax_c P(c) \prod_{i=1}^{n} P(f_i|c) = argmax_c P(c) \prod_{i=1}^{n} \frac{N_{ci} + \alpha}{N_c + \alpha n}.$$

where N_{ci} is the number of occurrences of the feature f_i in category c, and N_c is the total number of occurrences of every feature in category c. This paper uses Laplace smoothing in which α is 1, and the top five predictions are provided at the IPC subclass level. In addition, the 10-fold cross validation method is used to divide the patent document collection into a training set and test set.

4 Experiment and Evaluation

4.1 Precision Measurements

The goal of this paper is to provide a practical patent classifier that is able to support human decision-making processes. Therefore, two precision measures – single match precision and all match precision – are used to evaluate the classification performance for each field of the patent documents. Single match precision considers the classification successful if one of the predicted IPC subclasses for a patent document corresponds to one of the true IPC subclasses of that patent document. All match precision considers the classification successful if all of the predicted results for a patent document correspond to the true subclasses.

$$Precision_{single} = \frac{1}{n}\sum_{i=1}^{n} I(|T_i \cap P_i|)$$

$$Precision_{all} = \frac{1}{n}\sum_{i=1}^{n} I(|T_i \subseteq P_i|)$$

where I is the indicator function, T_i is the true IPC subclass categories of patent document i, P_i is the predicted IPC subclass categories, and n is the number of patent documents in the test patent documents set.

4.2 Evaluation

To examine the effects of the technical field and background, comparative experi-ments are designed with: (1) titles, (2) abstracts, (3) claims, (4) technical fields and backgrounds (tB), (5) titles and abstracts (TA), (6) titles, abstracts and claims (TAC), (7) titles, abstracts, technical fields and backgrounds (TAtB), (8) titles, abstracts, claims, technical fields and backgrounds (TACtB), according to the used fields for indexing. The experiments are performed with an Intel Xeon 8-Core CPU, 128 GB RAM and 64-bit Linux OS. The multinomial naive Bayes classifier is implemented in Python using the scikit-learn library [14].

IPC classification is performed for the patent documents set consisting of these eight different combinations of fields. The results show the highest precision (87.2 %) for (7) TAtB, which means that classification is successful for 492,513 of the total 564,793 test documents. 86.60 % precision is achieved for (4) tB, and 85.67 % precision for (8) TACtB. However, the classification performance shows the lowest rate of precision (76.66 %) for (3) claims, 77.57 % precision for (2) abstracts, and 78.24 % precision for (1) titles. For the second precision measure – all match precision – (7) TAtB is determined to be 70 %, and (4) tB and (8) TACtB are determined to be 69 % and 68.31 %, respectively.

It can be concluded that technical field and background play important roles in patent classification at the subclass level of the IPC, while claims decreases the precision of patent classification.

Looking at the results closely, the patent ‘Battery separator, and battery separator manufacturing method’ has five IPC subclasses of H01 M (Processes or means, e.g. batteries, for the direct conversion of chemical energy into electrical energy), B32B (Layered products, i.e. products built-up of strata of flat or non-flat, e.g. cellular or honeycomb, form), C08 J (Working-up; General processes of compounding; After-treatment not covered by subclasses C08B, C08C, C08F, C08G or C08H), C08 K (Use of inorganic or non-macromolecular organic substances as compounding ingredients) and C08L (Compositions of macromolecular compounds). When this patent document is classified using (4) tB and (7) TAtB, the results are correctly predicted for the five

Table 1. Classification precision at IPC subclass level

Indexing fields	Precision (single)	Precision (all)
(1) Titles	78.24 %	60.78 %
(2) Abstracts	77.57 %	59.40 %
(3) Claims	76.66 %	58.65 %
(4) Technical Fields, Backgrounds	86.60 %	69.00 %
(5) Titles, Abstracts	79.59 %	61.62 %
(6) Titles, Abstracts, Claims	79.13 %	61.18 %
(7) Titles, Abstracts, Technical Fields, Backgrounds	87.20 %	70.00 %
(8) Titles, Abstracts, Claims, Technical Fields, Backgrounds	85.67 %	68.31 %

IPC subclasses. When using (2) abstracts, (5) TA and (8) TACtB, four of the five IPC subclasses are correctly predicted, with the exception being C08 K. When using (1) titles, (3) claims and (6) TAC, three IPC subclasses are correctly predicted, with the exceptions being C08 K and C08L (Table 1).

5 Conclusions

In order to classify patent documents effectively based on the IPC system, this paper focused on the data structure of the patent documents themselves rather than on the classification models. It was suggested that the technical fields and background of the patent documents were significant in the classification task at the IPC subclass level. This suggestion was derived from an analysis of the characteristics of the patent documents and the IPC structure. An experiment was conducted to determine the effects of the structural fields on the classification performance for 564,793 Korean patent documents at the IPC subclass level. As a result, the use of the title, abstract, technical field, and background achieved the highest classification precision. In every case that included two significant fields, technical field and background showed better performance than their counterparts did. However, the use of claims caused a decrease in classification precision at the IPC subclass level. It was confirmed that the classification of Korean patent documents could be applied in the real world through multi-label classification at the IPC subclass level, which is divided into 630 categories.

However, further research is needed to construct a complete multi-label IPC auto-classification system that targets the lowest IPC level, which contains over 65,000 subgroups. To accomplish this, future research could employ a hierarchical algorithm that uses each patent document field selectively in each classification process of IPC based on the characteristics of the patent documents.

Acknowledgments. This research was supported by Gyeonggi Province's GRRC Program [(GRRC-B01), Development of Ambient Mobile Broadcasting Service System].

References

1. Choi, D.K.: Intellectual Property Statistics for 2014, Korean Intellectual Property Office (2015)
2. Seneviratne, D., Geva, S., Zuccon, G., Ferraro, G., Chappell, T., Meireles, M.: A signature approach to patent classification. In: Zuccon, G., Geva, S., Joho, H., Scholer, F., Sun, A., Zhang, P. (eds.) AIRS 2015. LNCS, vol. 9460, pp. 413–419. Springer, Heidelberg (2015). doi:10.1007/978-3-319-28940-3_35
3. Kim, J.-H., Choi, K.-S.: Patent document categorization based on semantic structural information. Inf. Process. Manage. **43**(5), 1200–1215 (2007)
4. Larkey, L.S.: A patent search and classification system. In: The 4th ACM Conference on Digital Libraries, pp. 119–187. ACM (1999)
5. Fall, C.J., Törcsvári, A., Benzineb, K., Karetka, G.: Automated categorization in the international patent classification. ACM SIGIR Forum **37**(1), 10–25 (2003). ACM
6. Tikk, D., Biró, G., Törcsvári, A.: A hierarchical online classifier for patent categorization. In: Emerging Technologies of Text Mining: Techniques and Applications, pp. 244–267 (2007)
7. Chen, Y.-L., Chang, Y.-C.: A three-phase method for patent classification. Inf. Process. Manage. **48**(6), 1017–1030 (2012)
8. Park, C., Kim, K., Seong, D.: Automatic IPC classification for patent documents of convergence technology using KNN. J. KIIT **12**(3), 175–185 (2014)
9. International Patent Classification Guide. http://www.wipo.int/export/sites/www/classifications/ipc/en/guide/guide_ipc.pdf
10. KIPRIS (Korea Intellectual Property Rights Information Service) plus. http://plus.kipris.or.kr/
11. KLT2000, Korean Morphological Analyzer. http://nlp.kookmin.ac.kr/
12. Zhang, H.: The optimality of Naive Bayes. In: Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference. AAAI Press, Miami Beach (2004)
13. Kibriya, A.M., Frank, E., Pfahringer, B., Holmes, G.: Multinomial Naive Bayes for text categorization revisited. In: Webb, G.I., Yu, X. (eds.) AI 2004. LNCS (LNAI), vol. 3339, pp. 488–499. Springer, Heidelberg (2004). doi:10.1007/978-3-540-30549-1_43
14. Buitinck, L., Louppe, G., Blondel, M., et al.: API design for machine learning software: experiences from the scikit-learn project. In: ECML PKDD Workshop on Languages for Machine Learning (2013)

A Comparison of Data Mining Methods in Analyzing Educational Data

Euihyun Jung(✉)

Department of Computer Science, Anyang University,
602-14, Jungang-ro Buleun-myeon, Ganghwa-gun, Incheon, Korea
jung@anyang.ac.kr

Abstract. Although data mining has been considered as a silver bullet which magically extracts valuable information from the stacked and unused data, its too many methods frequently confuse and mislead researchers. Therefore, in order to get a satisfying result, researchers need plenty of experience to choose a proper data mining method suitable to the purpose of their research. Unfortunately, in the education field, there are a few studies to point out this problem. In order to resolve this issue, in this paper, a study was conducted to compare Neural Network, Logistic Regression, and Decision Tree on educational data from Korea Youth Panel Survey (KYPS). The result showed the prediction accuracies of the methods were meaningfully different, but it doesn't mean that the prediction accuracy is the only factor in decision of a specific method. Rather, the result suggested that researchers should consider various aspects of the methods to choose a specific method because each method has its own pros and cons.

Keywords: Data mining · Neural network · Logistic regression · Decision tree

1 Introduction

Since the vast amount of data has been explosively generated in every field of the human society, data mining has drawn researchers' attentions so much [1]. Data mining has been used in many application areas and has rapidly extended its territory [2]. Currently, its domain spans over most academic and industry areas such as finance, e-commerce, biology, sociology, etc. On the other hand, in the education field, data mining is relatively new and the researchers in the area recently began to apply it to their studies [3, 4].

In data mining, there are a lot of methods which can be chosen according to the type of data and the purpose of research [5]. To classify data, researchers can choose one from Decision Tree, Support Vector Machine, Bayesian Classifier, Neural Network, etc. If they want to find association rules, Apriori algorithm or FP tree algorithm can be selected. In order to group similar items from data, K-means or network analysis can be the best choice. Besides, although there are plenty of existing methods already, newer machine learning methods have risen recently in data mining world [6].

Basically, this variety is useful in many situations, but it can also easily overwhelm and confuse researchers. Especially, in the education field, since the studies based on

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_28

data mining are relatively rare, the best practice how to select a method for the educational data doesn't be widely known yet. Actually, there have been studies in applying individual data mining method to educational data, but the research of overall effects for each method in analyzing the data has been hardly conducted.

In this paper, to give a guide for the decision of data mining methods in the education field, we conducted the comparison of three data mining methods; Neural Network [7], Logistic Regression [8], and Decision Tree [9]. The methods are widely used in prediction and we applied them to analyze the educational data from Korea Youth Panel Survey (KYPS) [10]. The result showed the prediction accuracies of the methods were meaningfully different, but it doesn't mean that the prediction accuracy is the only factor in decision of a specific method. Rather, the result suggested that researchers should consider various aspects of the methods to choose a specific method because each method has its own pros and cons.

The rest of this paper is organized as follows. Section 2 describes the data set used in the analysis, the compared data mining methods, and the result of the analysis. In Sect. 3, the pros and cons of the methods are discussed and Sect. 4 concludes the paper.

2 Analysis

2.1 Data Set

In this paper, the data of the 2nd grade students of middle school in Korea Youth Panel Survey (KYPS) [10] was used. To make KYPS, the 2nd grade students of middle school in Korea were sampled and the sampled students and their parents had been interviewed for five years since 2003. In the first year, the total number of samples was 3,449. In the sample, male students were 1,725 and female students were 1,724. The variables used in the analysis are shown in Table 1.

The purpose of the analysis is to predict the computer entertainment behavior of the students in KYPS data with the three data mining methods. In order to perform the Neural Network analysis, the continuous variables (e.g., computer use time, study time) were recoded to locate between 0 and 1 by standardizing. Categorical variables with orderly meaning (e.g., self-control, school adaptation) were defined between 0 and 1 and categorical variables without orderly meaning (e.g., gender, participation in cyber-club) were converted to dummy valuables (1 = true (apply to); 0 = false (not apply to)).

In order to conduct Logistic Regression, some continuous variables including Neural Network analysis were converted to dummy variables by dividing several sections. As a result, Logistic Regression and Neural Network were run with 63 variables expanded from the original 35 variables. For the analysis with Decision Tree, the values of 1 and 2 in 5-Likert scale were converted to categorical variable as 'low'; the value of 3 was coded as 'average'; the values of 4 and 5 were categorized as 'high'.

For Neural Network and Logistic Regression, the dependent or output measure used in this study was an indicator of whether a student often pursuit (1) or do not pursuit (0) the computer entertainment behavior. For Decision Tree, the dependent or

Table 1. The input data selected from KYPS for the analysis.

Area	Input data
Personal (16)	Gender, self-control, aggression, self-esteem, stress, optimistic tendency, self-awareness of a trouble maker, life satisfaction, physical condition, psychological state, participation in cyber-club, hope for education, worry of others criticism, computer use time, leisure time, self-reliance
Family (9)	Parental expectation of children study, parent attachment, parent abuse, spouse violence, relation of sibling, domestic economy, educational background of father, family composition, the average monthly family income
School (10)	Teacher attachment, school adaptation, grade, worry of study, peer attachment, frequency of meeting friends, close friends number, deviant peers number, private education time, study time

output variable was categorized as three values as follows: the values of 1 and 2 were set (and interpreted) as 'low', the value of 3 as 'normal', and the values between 4 and 5 as 'high'.

2.2 The Data Mining Method

In order to predict the computer entertainment behavior, the three methods were chosen; Neural Network, Logistic Regression and Decision Tree.

Neural Network. There are several models in Neural Network. Among them, the Multi-Layer Perceptron (MLP) model was selected in the paper. The MLP consists of multiple layers of perceptron and it utilizes a supervised learning technique called backpropagation for training the network. To use the MLP, the number of hidden layers and the number of nodes in each hidden layers had to be decided. The model used two hidden layers. Nine nodes were located in the first hidden layer and eight nodes were located in the second hidden layer.

Logistic Regression. Similar to Linear Regression, Logistic Regression measures the relationship between the dependent variable and several independent variables. However, its dependent variable has to be categorical data, so the data result is categorized and it can be considered as a classification method. It uses the binary logistic model to get the probabilities of a binary response of input variables. Before using Logistic Regression, it is important to check multicollinearity and we made it sure that there was no multicollinearity among input variables.

Decision Tree. Chi-square Automatic Interaction Detector (CHAID) [11] algorithm was selected for the Decision Tree method in this paper. From the root node of a decision tree, CHAID uses the χ^2 test to check a null hypothesis which assumes that the frequency of the real occurrence and the expected occurrence is not different. When the null hypothesis is rejected, a new branch is flipped based on the input variables. The tree keeps evolving until its depth reaches a maximum tree depth defined as 5 in this paper. The p-value in the test was set under .05.

2.3 Results

In the comparison of the prediction accuracy, the overall Correct Classification Rate (CCR) of each method was measured. As shown in Table 2, Neural Network had the best CCR value and Decision Tree had the worst CCR value. The results suggested Neural Network should be selected when the prediction accuracy is most important factor. On the other hand, Decision Tree should be avoided in terms of the prediction accuracy.

Table 2. The evaluated CCR values of the three methods.

Method	CCR
Neural network	83.6 %
Logistic regression	79.8 %
Decision tree	62.5 %

3 Discussion

Neural Network achieved the highest degree of the prediction accuracy in the result. That means if the purpose of research is to predict someone's level of the computer entertainment behavior, Neural Network can be the best method to mine the educational data. However, if researchers wonder which input variable most affects someone's computer entertainment behavior, Neural Network cannot give any clue for that question. It also requires great amount of time and plenty of data to train the network through trial and error.

Among the three methods, Decision Tree is most behind in the prediction accuracy. However, its strength is not giving accurate prediction but explaining the degree of the input variables' contribution. As shown in Fig. 1, the result of Decision Tree analysis can be represented as a reverse tree consisting of nodes which indicate the degree of

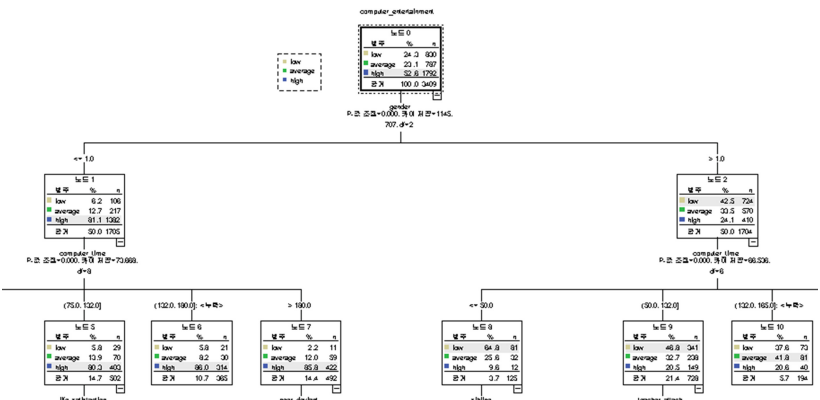


Fig. 1. The part of the result diagram from decision tree analysis.

contribution. By investigating the tree, researchers can easily figure out which input variable has most influence on the dependent variable. For example, in the analysis, researchers can suggest a solution to ease the computer entertainment behavior.

Logistic Regression is less accurate in prediction than Neural Network, but it does not require the training cost at all. Its simplicity is superior to other methods and the result is fair enough. Also, since it comes from traditional statistics, the usage of Logistic Regression has been well known. For example, with it, researchers can find which input variables affect the dependent variable and can get the probability of the dependent variable.

Regardless of the obvious difference of the prediction accuracy, the selection of the data mining method is not simple because the compared methods have their own pros and cons. Shortly, researchers should choose the method carefully according to the type of data and the purpose of research. The pros and cons of the data mining methods are summarized in Table 3.

Table 3. The pros and cons of the data mining methods in comparison.

Method	Pros	Cons
Neural network	High prediction accuracy	Uncertainty of the effects from input variables
		Needs of computational cost, enough time, and plentiful data for training the network
Logistic regression	Handy and simple computation	Relatively low prediction accuracy
	Getting additional information	
Decision tree	Explaining the degree of individual input variables' contribution	Low prediction accuracy

4 Conclusion

Data mining is considered as a silver bullet that magically extracts valuable information from the stacked data in various fields, but it has a lot of methods to confuse and mislead researchers. In order to get a satisfying result, researchers should have plenty of experience to choose a proper data mining method suitable to the type of data and the purpose of the research. Unfortunately, in the education field, there are a few studies to point out this problem.

In this paper, to give a guide for the choice, we conducted the comparison of three data mining methods which are widely used in prediction. In the result, each method showed its own pros and cons. Neural Network was best at the prediction, Logistic Regression was handy and fast and Decision Tree had the advantage in explaining the degree of input variables' contribution. The result means there is no single method to fit every situation at all and it is important to select a proper method for the purpose of research. In the future, we are going to compare more data mining methods including unsupervised learning for data in the education field.

References

1. Larose, D.T.: Discovering Knowledge in Data: An Introduction to Data Mining. Wiley, New York (2014)
2. Wu, X., Zhu, X., Wu, G.Q., Ding, W.: Data Mining with Big Data. IEEE Trans. Knowl. Data Eng. **26**(1), 97–107 (2014)
3. Romero, C., Ventura, S.: Data mining in education. Wiley Interdisc. Rev. Data Mining Knowl. Discov. **3**(1), 12–27 (2013)
4. Baker, R.S., Inventado, P.S.: Educational data mining and learning analytics. In: Larusson, J.A., White, B. (eds.) learning analytics, pp. 61–75. Springer, New York (2014)
5. Aggarwal, C.C.: Data Mining: The Textbook. Springer, New York (2015)
6. Dean, J.: Big Data, Data Mining, and Machine Learning: Value Creation for Business Leaders and Practitioners. Wiley, New York (2014)
7. Hagan, M.T., Demuth, H.B., Beale, M.H., De Jesús, O.: Neural Network Design, vol. 20. PWS Publishing Company, Boston (1996)
8. Hosmer Jr., D.W., Lemeshow, S.: Applied Logistic Regression. Wiley, New York (2004)
9. Geatz, M.W., Roiger, R.: Data Mining: A Tutorial Based Primer. Addison Wesley, Boston (2003)
10. National Youth Policy Institute: Korea Youth Panel Survey (KYPS) User Guide, Seoul (2010)
11. Kass, G.V.: An exploratory technique for investigating large quantities of categorical data. Appl. Stat. **29**, 119–127 (1980)

A New Secure Android Model Based on Privilege

Tao Zhang(✉) and Zhilong Wang

College of Information Science and Engineering, Ocean University of China,
No. 238 Songling Road, Qingdao, Shandong, China
oucseczt@l26.com, zhilongwang@foxmail.com

Abstract. Android is the most popular smartphone operating system in the world. There are many people who focus on Android security and dedicated to improve android security. There are many android vulnerabilities exposed online and attackers can use these vulnerabilities to steal our private and sensitive information and attack our device. In the paper, we propose a novel secure android model based on privilege. We create three kinds of users and grant different users different permissions. Thus, users can have more freedom and control over their android device in the model we proposed. In our secure model, users can upgrade their android operating system in time, which may enhance the security of their smartphones and protect users' sensitive information better.

Keywords: Android · Security model · Root · Privilege

1 Introduction

1.1 Android Introduction

Led by Google, Android is an open source operating system which based on the Linux kernel. Android is the first open and free mobile operating system that is used widely on the mobile devices.

Google is trying to ensure that Android is a safe and open ecosystem for users and developers [1]. Android is the most popular smartphone operating system in the world. There are about 81.6 percent android device and 15.9 percent iOS devices in the worldwide smartphone operating market share in 2015 [2]. There are more than 1.4 million apps in total dating back to Dec 31, 2015. According to Yahoo Flurry, there is a 14 percent increase in the android apps, and a 332 % boost in some special area. Users can download and install applications from Google play and other third party application markets. However, there are some malicious applications and applications with vulnerabilities in these application markets.

1.2 Android Security Status

As more and more people use smart phones, phones are not just phones. We can use smart phones to chat with others by WeChat, MSN and other applications, to record our

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_29

life by taking photos and to do shopping online. Smartphones store too much sensitive information and personal information, such as contacts, SMS, bank accounts, payment information, photos, GPS information. In a word, security for smartphone is becoming an issue [3].

Android is open source and the AOSP (Android Open Source Project) get broad security review by many interested parties, including hackers, manufacturers and other interested parties. There are many android vulnerabilities exposed online in the past years and attackers can use these vulnerabilities to steal our private and sensitive information, even to attack our android devices. Of these vulnerabilities, some are fatal and can bring great damages to devices and the owner. Take stagefright as example, CVE-2015-1538 [4], it can be abused to control our mobile phones by privilege escalation.

There are many different android versions customized by different customers and it takes months for users to get access to the version of system security upgrade or update. What is worse, there are too many users who have no idea to update their android operating system.

2 Android Security Mechanisms

There are many security mechanisms in android to ensure the security of the most popular mobile operating system, such as permission model, sandboxing and isolation.

2.1 Permission Model

Permission model is an application level security mechanism of android. There is a *AndroidMainfest.xml* file that describes the permission needed during execution in a application's APK package. One can set *uses-permission* to request permission when developing an application.

At runtime, services at this layer enforce the permissions specified in the manifest and granted by the user during installation [5].

Permission and permission enforcement make sure that an application's access to resource is in accord with the description in *AndroidMainfest.xml* file.

2.2 Sandboxing and Isolation

Sandboxing and isolation are kernel level security mechanism of android. All android devices share a common security model that provides every application with a secure, isolated environment known as an application sandbox [1]. Android security model is partly based on application sandboxes. Each application is executed in a separate Dalvik VM machine, which ensures isolation among applications [5].

2.3 Security Enhanced Android

SELinux, means Security Enhancements for Linux, is mandatory access control for Linux, which can confine flawed and malicious applications and prevent privilege escalation.

SE android [6], Security Enhancements for android, is available in Android since version 4.4.3, which enforce SELinux in android to mitigate malicious applications' threat and to prevent privilege escalation, data leakage by applications and bypass of security features.

2.4 Root

Root is the privilege user of android and Linux operating system. Actually, we don't have full control over our android devices [7]. A user or application with root permissions can modify the operating system, kernel files and any other applications' permissions. If we have full control over the device, we can change system settings, uninstall system applications, delete system files, backup system application data, install applications at SD card and so on.

3 Android Security Framework Analysis

3.1 Permission Model

Although many security mechanisms applied in the android operating system, both on Linux kernel layer and application layer. There are many vulnerabilities in android which may be abused and make our private data in danger. Android enforces strict security mechanisms to limit code execution of vulnerabilities, which may introduce new vulnerabilities at the same time. In android, some malicious applications can bypass those security mechanisms by exploiting some vulnerabilities [8].

However, the permission model is all or nothing [9]. Actually, the user need to accept all permission the application required or the application can not be installed successfully. Once successfully installed, the application will permanently have access to the permission described in the *AndroidMainfest.xml* file. However, for example, if an application requires the GPS permission when installed, it can have access to GPS information whenever. Users do not have flexible control over the permission the application required and their sensitive information, such as GPS information.

Permission and permission enforcement make sure that a application's access to resource is in accord with the description in *AndroidMainfest.xml* file. However, the permission model is not as secure as we think. In [10], the author modify the *AndroidMainfest.xml* file, then reboot the smartphone and the application gets the permission he modified before.

3.2 SE Android

There are also some limits in SE android. SE android is SE-Linux's migration and extension in android, which focus on enforcement access to system call and kernel. SE-android provide solution to malicious applications' abuse system call to invoke kernel function in android [5]. But there is no protection on user data which may be stolen and modified [11] and there is not existing plans to mitigate common vulnerabilities [10].

3.3 Root

As a privilege user, root has full access to all applications and all data on the device. Basebridge malware can use HTTP protocol to communicate with central server and convey our private information. DroidKongFu can collect IMEI, device ID, SDK versions and other information which will be sent to a specific remote server [12]. Androidkungfu use *NPROC_RLIMIT* vulnerability and FramaRoot that were reported before to acquire root privilege, then control the victims' smartphones [8]. Root access gained by exploiting a kernel bug or security hole can bypass bootloader unlock mechanism protection. Once malicious applications or applications with vulnerabilities run with root privilege, all your information on the devices will never be yours.

Most of these vulnerabilities exploit vulnerabilities to gain root access or escalate privilege, which are among the largest threats to android operating system.

4 A New Android Security Framework

4.1 Root-Admin-User

There should be a balance between the security of android and users' flexible control over smartphones. Android users want to get more control over their android devices, so some of them try to root their devices and there are many kinds of rooting tools available. However, if we can not protect the root privilege properly, which may cause great damage to android devices and their owners. And there are many malicious applications which use vulnerabilities to get root privilege or escalate privilege.

Like Linux user model, we propose a new security model and divide users into 3 kinds: root, admin and user.

Root: In our model, when users get their android device, they should not get root privilege again unless they format or upgrade their smartphones. We will grant most of the root privilege to the admin and normal user, for example, we give normal user the privilege to change fonts of their android operating system. To assure the security of the android operating system, some of the key system files are not allowed to be changed in our model, such as the *.lib* file.

Admin: We assign the admin user most of the privileges which are belonged to root before. The android operating system run as normal user initially. If normal user need to get root privilege, it will switch to get admin privilege in our model.

User: User is the ordinary android user with no privilege in our model. However, we give owner of the android device more control over the device as a normal user. For example, the owner of the device can uninstall the pre-installed applications.

4.2 Classify System File and Rearrange Privilege

We are to give users more freedom to control and set their android devices. What's more, we provide a more secure and privacy-protected framework of android, so we can protect user data and privacy better on these android devices.

We divide system files into 3 parts based on privilege, then enforce the access control mechanism.

The 1st part is the key files that should not be changed. The 2nd part is the files that can be changed by users freely, and the 3rd is the rest files. We give these 3 kinds of files different privileges, the 1st kind of files is owned by root, which should not be changed in the security model we proposed. The 3rd is owned by admin that can be changed intentionally and users can access and change these files only in some condition. The 2nd kind of files is those files owned by user and the owner of the device can change these files freely, such as fonts, background.

4.3 Upgrade Mechanism

Android is open source and allow device manufacturers to customize devices and introduce diversity. In the android security ecosystem, Google is far away from end users [1], because the android system running on our devices is accustomed by device makers and we can not improve our android's security although Google released the security improvements.

Android is an open operating system and openness will strengthen security [1]. All interested parties can join to improve android's security levels. But there should be some organizations like Google or android union who be responsible for the upgrade of the core system.

The upgrade of operating system is responsible for different parties, such as device manufacturers, vendors and an organization like Google and Android Security Unions. End users can upgrade their android directly after the organization released security upgrade, even their device customers did not provide such security upgrade.

The device customers' upgrade is about the customized part of system and this kinds of upgrade does not have an effect on the upgrade of kernel.

5 Conclusions

In the paper, we propose a more secure and privacy-protected android model which changes the permission model so that users can have better control over their android devices and make sure their data and private information get better protection. We create three kinds of users and grant different users different privilege. In our secure model, users can upgrade their android operating system in time, which will enhance the security of their smartphones and privacy.

References

1. Google. Google android security 2015 report. http://static.googleusercontent.com/media/source.android.com/en//security/reports/GoogleAndroid. Security 2015 Report Final.pdf
2. Statista. Android os market share of smartphone sales to end users from 2009 to 2015. http://www.statista.com/statistics/216420/
3. Han, K.S., Lee, Y., Jiang, B., Im, E.G.: Android permission system violation: case study and renement. Int. J. E-Entrepreneurship Innov. (IJEEI) **4**(1), 16 (2013)
4. The Latest on Stagefright: CVE-2015-1538 Exploit is Now Available for Testing Purposes. https://blog.zimperium.com/the-latest-on-stagefright-cve-2015-1538-exploit-is-now-available-for-testing-purposes/
5. Merlo, A., Costa, G., Verderame, L., et al.: Android vs. SEAndroid: an empirical assessment. Pervasive Mob. Comput. **30**, 113–131 (2016)
6. Smalley, S.: The Case for SE Android. National Security Agency (2011)
7. Zhang, H., She, D., Qian, Z.: Android root and its providers: a double-edged sword. In: ACM Sigsac Conference on Computer and Communications Security, pp. 1093–1104. ACM (2015)
8. University of Chinese Academy of Sciences National Computer Network Emergency Response Technical Team Coordination Center of China State Key Laboratory of Integrated Services Networks Xidian University Beijing Electronic Science and Technology Institute. Survey of Android Vulnerability Detection. J. Comput. Res. Dev. (2015)
9. Nauman, M., Khan, S., Zhang, X.: Apex: extending android permission model and enforcement with user-defined runtime constraints. In: ACM Symposium on Information, Computer and Communications Security, ASIACCS 2010, Beijing, China, April 2010
10. Yang, C., et al.: Utilization pattern based android root vulnerability analysis. Comput. Sci. **41**, 343–346 (2014)
11. Xiang, W.H., Yu-Jun, L.I., Hou, M.S.: Research on privacy protection based on SEAndroid. Comput. Sci. **42**, 329–332 (2015)
12. Zhou, Y., Jiang, X.: Dissecting Android malware: characterization and evolution. In: IEEE Symposium on Security & Privacy. IEEE (2012)

Survey of MCC Architectures for Computing Service

Byeong-Seok Park, Yoon-A Heo, and Young-Sik Jeong(✉)

Department of Multimedia Engineering,
Dongguk University, Seoul, South Korea
{ggoommy, hyagood, ysjeong}@dongguk.edu

Abstract. Following the growth of mobile device applications in the past years, mobile computing and cloud computing are suggesting a new computing paradigm. Demands for computing capability regarding potential technologies that are provided as mobile service are also growing. Meanwhile, necessity of Mobile cloud computing (MCC) is increasing due to the problems of mobile devices, such as resource poverty, battery constraint. However, clear definition is required as the concept of MCC is still ambiguous. Currently, it is necessary to classify the configuration methods of the resource-providing infrastructure environment of MCC. This paper provides a comprehensive overview of MCC and proposes a basic scheme. We suggest instance MCC (iMCC), which is a framework implementing offloading for the purpose of achieving efficient computing process and service provision. iMCC includes improved resource allocation method in MCC infrastructure as well as efficient offloading method. Finally, we summarize the conclusion and discuss future research.

Keywords: Mobile cloud computing · Architecture · Offloading

1 Introduction

Recently, mobile devices, from smart phone to wearable devices such as smart watch, are playing a central role in the actualization of mobile convergence in the ubiquitous society as fast and convenient communication tools. Most people have mobile devices and their utilization as information processing tools are increasing and getting more pervasive. The applications of mobile devices are rapidly growing in video games, augmented reality, and wearable computing as well as in the traditional categories such as entertainment, health, and social networks. As a result, people are demanding greater computing capability. Mobile devices, however, still have constrained resources, limited battery, and low bandwidth compared to traditional workstations. To address this problem, mobile devices can get resources from external source such as cloud computing platform. In particular, mobile cloud computing (MCC) which provide resources from a remote server to receive computing and storage services can overcome many issues of mobile devices. In this context, the demand for MCC which provides users with required tools and services anywhere, anytime is continuously increasing with the explosive growth of mobile applications and cloud computing. The importance of MCC is gradually increasing and related research is being actively conducted. Due

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_30

to the broad concept of MCC, the components of MCC are not clearly defined yet, and the construction methods of the resource-providing infrastructure environment of MCC are not clearly classified. MCC has the possibility of overcoming the device performance limitations such as battery life, storage, and bandwidth, as well as the environment issues such as heterogeneity, scalability, and availability while preserving the nature of mobile devices. Nevertheless, the potential of MCC has not been fully realized because the components and architecture of MCC are not standardized and different concepts are mixed.

This paper provides a comprehensive overview of MCC and proposes a basic scheme. Section 2 defines the concept and components of MCC in various domains and introduces related works. Section 3 describes the MCC architecture and classifies MCC frameworks that have been researched until now are classified according to the cloud infrastructure construction method. Section 4 proposes a cloud computing model to provide more efficient computing service from the aspect of performance in specific situations. This model includes improve resource allocation, efficient task distribution, and offload in the MCC infrastructure environment. Finally, Sect. 5 summarizes the conclusion and discusses future research subjects.

2 Related Works

MCC Forum defines MCC as 'an infrastructure where both the data storage and data processing happen outside of the mobile device [1]. MCC applications use computing power and data storage from the outside sources that have rich resources through a cloud. MCC does not just mean the expansion of cloud computing concept to the mobile, but it also means that people can use desired services with no limitations in time or space using cloud technology on any mobile device including smart phones that can communicate. In a cloud environment, mobile devices overcome the aforementioned potential issues of MCC through offloading or dynamic configuration of services. It can be said to be an Application-as-a-Service which performs applications by providing computing service through a mobile device based on an existing cloud service.

Le Guan [2] classified the architectures of MCC into the following two schemes: agent-client scheme and collaborated scheme. The agent-client scheme overcomes the limited processing power and data storage of mobile devices by managing all resources of mobile devices by cloud. In other words, this is a centralized scheme where the cloud communicates with and manages devices by creating an agent on each device. In the collaborated scheme, the cloud server takes charge of the functions of controller and scheduler for collaboration among mobile devices.

Dipayan Dev [3] suggested a structure where mobile devices are connected to the cloud through wireless access point and radio tower and the cloud access remote servers and virtual machines. A cloud server with a security mechanism through firewall is deployed between the cloud and the remote server.

3 MCC Architecture

In this chapter, the architecture of MCC for the organization of MCC and the provision of computing services is examined. In general, the concept of MCC focuses on enhancing the insufficient computing capability of mobile devices. Meanwhile, there is another angle of MCC that it is realized to use the idle resources of other mobile devices. In this paper, considering these two concepts, the MCC architecture schemes are classified into the following two types: Server agent scheme and Inter-mobile scheme.

3.1 Server Agent Scheme

The server agent scheme which is a general MCC is shown in Fig. 1. In this figure, mobile devices are connected to the internet by mobile network through satellite (3G, 4G) or wireless access point (Wi-Fi). The mobile networks provide mobile users with such services as authentication. Users access the cloud services provided by cloud service providers through the Internet connection. The mobile cloud middleware of the cloud handles the cloud service requests of mobile users. The Virtual Machine (VM) configures a resource pool for integrating and managing application servers that provide computing power and data centers that provide storage. The service requests of mobile users are carried out through distributed computing with the available resources that belong to multiple clouds. The server agent scheme works in this architecture based on the concepts of utility computing, virtualization, and service-oriented architecture.

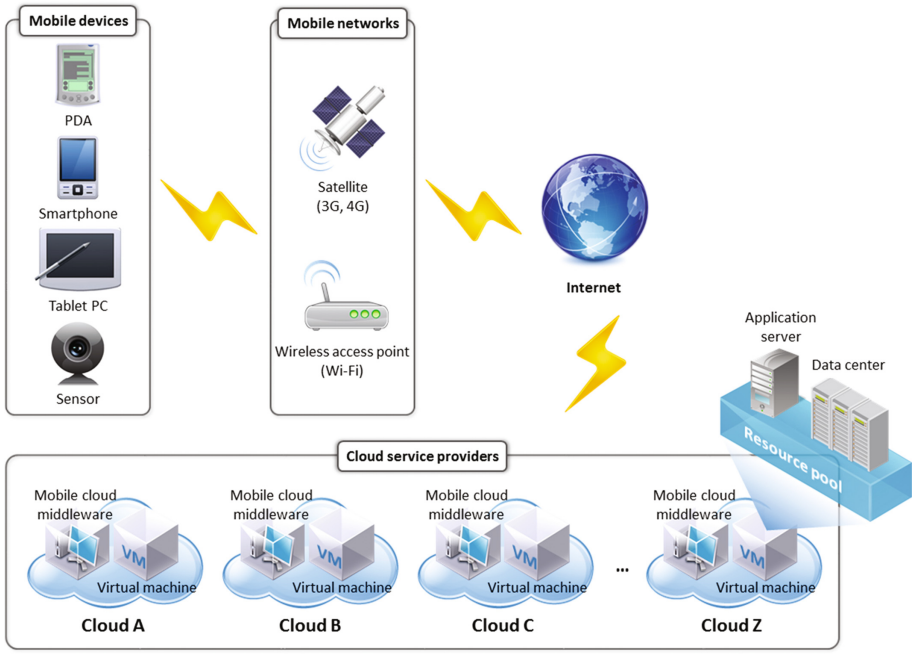


Fig. 1. Server agent scheme

Cloudlet [4] is a solution that has been suggested instead of connecting to a remote cloud considering that the performance of MCC is considerably lowered when the mobile device has difficulty connecting with the remote server. The cloudlet concept is similar to a small data center. It serves as an intermediate data center for connecting to a large cloud with richer resources through the Internet. It is ideal for specific local businesses such as coffee shops because it is appropriate in a specific area or place.

3.2 Inter-mobile Agent Scheme

Unlike the server agent scheme which receives computing services from remote servers through the mobile network, this scheme regards other mobile devices as part of the cloud and uses their idle computing capability as resources. Mobile devices integrate resources by communicating with others through Cloud controller. Resource manager forms a pool by integrating the resources of mobile devices and receives computing power to handle the distributed tasks. This scheme has an advantage that can use idle resources from the neighborhood when network to the remote server is not stable. However, this scheme requires the configuration of an environment where the scheme can be used and the issue of how to provide incentives to mobile devices participating as service providers of a cloud still remains. Furthermore, it is difficult to predict the number

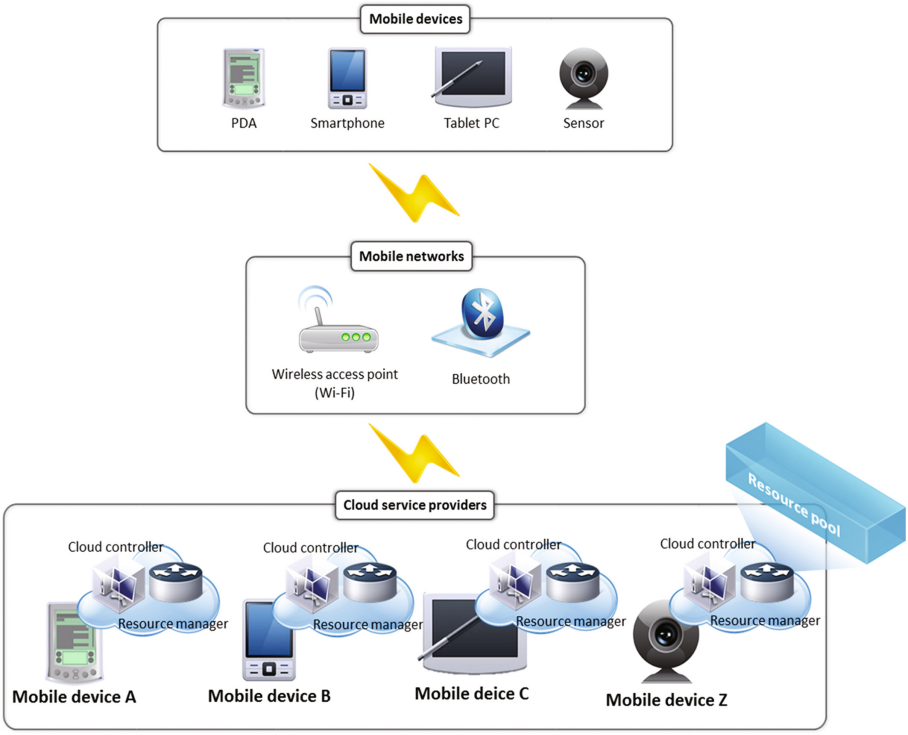


Fig. 2. Inter-mobile agent scheme

and types of available resources in advance and the method of efficient task distribution and road balancing among the mobile devices needs to be considered (Fig. 2).

Hyrax [5] is an Android application that researched the usability of clusters of mobile devices as resource provider. The devices adjust data and tasks by communicating through an independent 802.11g network. Each mobile device runs a thread on the HDFS to save the multimedia data of the device and record sensor data. As a sample application, HyraxTube was presented which allows users to search multimedia files by time, quality, and location.

4 instant Mobile Cloud Computing (IMCC)

This paper proposes instant Mobile Cloud Computing (iMCC) which provides a cloud computing infrastructure platform using mobile devices. This model regards mobile devices that are geographically close as resource providers to distribute the tasks of a single device to multiple devices. The architecture of iMCC that works in the inter-mobile agent scheme is shown in Fig. 3.

In the iMCC model, users participate in the cloud through the iMCC application. The iMCC configures a cloud by forming an Intranet through the Wi-Fi. The CPU, D RAM, Flash Memory, and GPU performances of every mobile device are resource-pooled and form pools such as Storage Pool, Network Pool, CPU Pool, and Memory Pool. The inputs coming from the iMCC application are load-balanced by the

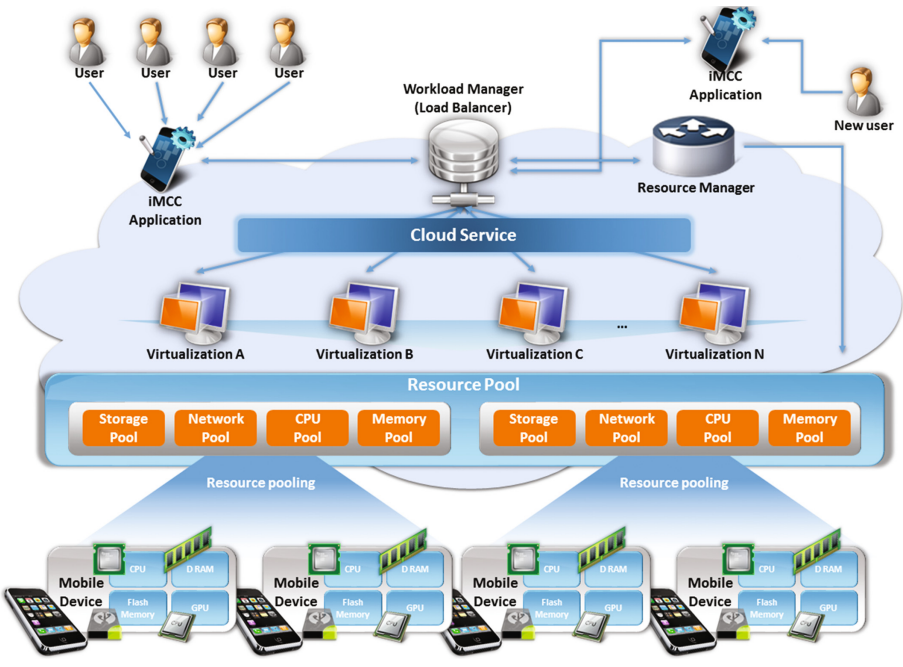


Fig. 3. iMCC architecture

Workload Manager and the tasks are distributed in the virtual resource pool. In the iMCC, mobile devices can more freely participate as or withdraw from cloud resource providers than in the server agent scheme. Therefore, the Resource Manager continuously monitors and manages mobile resources. The iMCC is especially useful for providing cloud computing services through mobile devices on an airplane or ship where network connection is difficult.

5 Conclusion

This paper proposed two schemes of MCC architectures from the aspect of cloud infrastructure configuration to provide computing services based on existing studies. Furthermore, the MCC frameworks that have been studied in the past were analyzed to examine in which one of the two schemes they work. Furthermore, the iMCC model was proposed which is appropriate for implementing the MCC in an environment where Internet connection is difficult. The iMCC organizes a cloud with surrounding mobile devices through Wi-Fi in a mobile network environment where it is difficult to access the remote server through a satellite, and it can be used even with a small network bandwidth and intermittent interruptions. Based on the results of this study, the iMCC model will be researched further in the future.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2053564). And also This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP (Institute for Information & communications Technology Promotion), And also This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R7120-16-1006, Resource Integration Management Solution with On-premises Legacy for Intra-Cloud).

References

1. MCC Forum. http://www.mobilecloudcomputingforum.com
2. Guan, L., Ke, X., Song, M., Song, J.: A survey of research on mobile cloud computing. In: Proceedings of the 10th IEEE/ACIS International Conference on Computer and Information Science. IEEE Computer Society, pp. 387–392 (2011)
3. Dev, D., Baishnab, K.L.: A review and research towards mobile cloud computing. In: The Proceedings of the Second IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, pp. 252–256 (2014)
4. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The case for VM-Based cloudlets in mobile computing. IEEE Pervasive Comput. **8**, 14–23 (2009)
5. Marinelli, E.E.: Hyrax: cloud computing on mobile devices using MapReduce, Masters thesis, Carnegie Mellon University (2009)

6. Huerta-Canepa, G., Lee, D.: A virtual cloud computing provider for mobile devices. In: Proceedings of the 1st ACM Workshop on Mobile Cloud Computing & Services: Social Networks and Beyond, MCS 2010, pp. 6:1–6:5. ACM, New York (2010)
7. Fernando, N., Loke, S.W., Rahayu, W.: Mobile cloud computing: a survey. Future Gener. Comput. Syst. **29**(1), 84–106 (2013)

Measurement of Enterprise Smart Business Capability in a Global Management Environment

Chui Young Yoon(✉)

Department of IT Applied-Convergence, Korea National University of Transportation, Chungju, Chungbuk 368-701, Republic of Korea
yoon0l09@ut.ac.kr

Abstract. Enterprises are endeavoring to effectively apply smart business technology to their management activities in order to raise their business results in a smart business environment. Enterprise's smart business capability is very critical for the efficient execution of its management activities and to improve the performance of business tasks in a global management environment. An efficient measurement framework is necessary for efficiently measuring a firm's smart business capability to manage and improve its smart business capability. The developed 12-item scale was verified based on previous literature. We found a 12-item framework that can reasonably gauge an enterprise smart business capability. This framework can be used for efficiently measuring a firm's smart business capability in a comprehensive perspective.

Keywords: Smart business · Smart Business Capability · Measurement items · Measurement framework

1 Introduction

Many enterprises are performing their management activities and business tasks with partially and fully utilizing smart device, network, solutions and systems in a smart business environment [1–4]. Enterprises have built smart business environment to increase their task performance and to improve their competitiveness in a global management environment. Firms are also applying smart technology to the management and business activities such as smart business workstation, smart business mobile platform, and smart business solutions etc. Smart business technology is a critical method to improve a firm's management performance and competitiveness in the ever-changing business environment. It is indispensable to apply smart business technology to the management activities and business tasks of an enterprise. Enterprise smart business capability means the entire capability that a firm utilizes smart business technology for its management activities and for improves its business performances in a global management environment. Enterprise smart business capability has to be gauged by a scientific and practical measurement framework and should be improved by objective criteria based on the measurement results of the measurement framework. But a reasonable framework to measure a firm smart business capability has not been studied in

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_31

previous literature. That is, we need a reasonable framework that can efficiently gauge an enterprise smart business capability in terms of entire smart business ability.

Therefore, this study presents an efficient framework that can measure an enterprise smart business capability for effectively performing its smart business tasks and for properly improving its smart business performances in terms of a whole smart business capability.

2 Previous Research

Smart business can be described as a business process that uses the smart technology medium as a conduit to fulfill business transactions. Smart business can be defined as an approach to raise the competitiveness of organizations by improving management activities through using smart technology such as smart devices, networks, and solutions [1–4]. Smart business can be presented as a method to efficiently perform the firm's management activities by applying the smart business strategy, technology and solutions, and systems to its business tasks in a global management environment.

Many studies defined the concepts of IT capability from the view points of the studies' researchers [5–8]. But smart business capability has rarely researched in previous literature. IT capability is conceptualized as the extent to which a firm is knowledgeable about and effectively utilizes IT technology to manage IT data within the firm [9–11]. The components of IT capability represent three co-specialized resources: IT objects; IT knowledge; and IT operations [9–11]. IT capability is considered the culmination of the sets of hardware, software, services, management practices, and technologies and management skills related to IT departments [12]. IT capability is formed by IT system convention, IT infrastructure, human IT resources, and IT relationship assets based on these resource-based perspectives [13]. From an information system maturity system perspective, the measurement of the information system level indicates the total capability that includes information system vision, information system infrastructure, information system support, and information system application and usage [6, 14–16].

Therefore, this study defines the enterprise smart business capability (ESC) as the entire smart business capabilities that an enterprise has to retain to efficiently perform its smart business tasks and improve its smart business performances in a global management environment. This research develops the first measurement items for ESC based on the definition of ESC and previous studies related to an enterprise smart business capability.

3 Methods

From exploring the previous literature, we generated an initial list of 22 measurement items for ESC based on definitions and components of IT capability [9–16]. This research analyzed the validity and reliability of the developed items to ensure that ESC is efficiently measured by the items. It was proved by presenting that the framework was a suitable operational definition of the construct it purported to measure.

Many studies presented various methods to verify the validation of a model construct [14–18]. Generally, most studies present two methods of construct validation: (1) correlations between total scores and item scores, and (2) factor analysis [14–18].

In this research, the measurement questionnaire used a five-point Likert-type scale as presented in previous studies; denoting, 1: not at all; 2: a little; 3: moderate; 4: good; and 5: very good. The survey was gathered data from a variety of industries, business departments, experience, and education. We performed two kinds of survey methods: direct collection and e-mail. The respondents either directly mailed back the completed questionnaires or research assistants collected them 2-3 weeks later. The collected questionnaires represented 43 % of the respondents.

3.1 Sample Characteristics

In this questionnaire survey, this research collected 114 responses form 265 respondents. They represented a variety of industries, enterprises, business departments and positions, and experience. We excluded two incomplete or ambiguous questionnaires, leaving 112 usable questionnaires for statistical analysis. All respondents had college or university degrees in: humanities and societies (17.1 %), management and economics (33.3 %), engineering (30.7 %), and science (18.9 %). The respondents in terms of business departments were identified as strategy planning (17.5 %), development and maintenance (16.7 %), business application (34.2 %), and administration support (31.6 %). The respondents had on average 8.9 years' experience (S.D. = 1.03) in their field, their average age was 36.5 years old (S.D. = 5.97), and 68 % were male. This survey was intentionally focused on various industries and persons working above the 5 years within their firms. Namely, the respondents could efficiently provide the correct responses for our questionnaire survey.

3.2 Analysis and Discussions

This research extracted the various analysis results from the collected usable questionnaires. After factor analysis and reliability analysis, the first 22 measurement items were reduced to 12 items, with 10 items were deleted, with applying the criterion of previous studies [16–18]. The elimination was sufficiently considered to ensure that the retained items were adequate analysis items of ESC. Each of the 12 items had a factor loading > 0.611. The reliability coefficients (Cronbach's alpha) of four potential factors had values > 0.789 as indicated in Table 1, above the threshold recommended for exploratory research [16–18]. We calculated the corrected item-total correlations between each variable and its corresponding factor in order to investigate the reliability and validity of the measurement items. We considered sufficiently high criteria to extract reasonable analysis items of ESC. These coefficients indicate the relative contribution of a measurement item for the construction of a scale to gauge a particular factor. Most corrected item-total correlations were greater than 0.602, showing that the measurement items are good indicators of their corresponding factors. The extracted items have validity and reliability in terms of a measurement construct based on the measurement results as indicated in Table 1.

Table 1. Reliability, validity, and factor loadings of ESC construct.

Variable	Factor loading				Corrected item-total correlation	Coefficients alpha
	Factor 1	Factor 2	Factor 3	Factor 4		
V02	0.781				0.625	
V03	0.774				0.698	0.789
V05	0.623				0.713	
V07		0.798			0.704	
V09		0.814			0.621	0.816
V10		0.739			0.701	
V12			0.807		0.638	
V14			0.813		0.681	0.809
V15			0.714		0.667	
V18				0.729	0.628	
V19				0.753	0.701	0.801
V21				0.611	0.602	

These results may be successfully achieved by accumulating many research findings and case studies. The developed measurement framework can be become more objective and practical scale in the application of industrial fields, with reflecting the measurement results of many case studies.

4 Measurement Framework of ESC

The extracted 12 items were classified into four factor groups based on the factor analysis. The factor groups indicate the potential factors as major components to measure ESC. We identified the following four potential factors by exploring the measurement items of each factor group based on previous studies: factor 1: smart business strategy; factor 2: smart business technology; factor 3: smart business utilization; and factor 4: smart business infrastructure. These extracted factors include the overall measurement content for ESC from smart business strategy to smart business infrastructure. As presented in Table 1 and Fig. 1, smart business strategy presents a firm's consistent smart business policy and plan such as smart business plan and program, consentaneity between smart business plan and management plan, and smart business trends, including V02, V03, and V05. Smart business technology indicates the technical knowledge that a firm has to retain such as smart business solutions and applications, and smart systems, with V07, V09, and V10. Smart business utilization represents a firm's ability to utilize smart business solutions, big data and cloud technology, and smart security solutions and systems for efficiently executing the firm's smart business tasks, containing V12, V14, and V15. Smart business infrastructure refers to smart business resources that a firm retains for smart business, such as smart business systems, intellectual property, and smart business security measures and systems, comprising V18, V19, and V21. Our findings present a framework that can

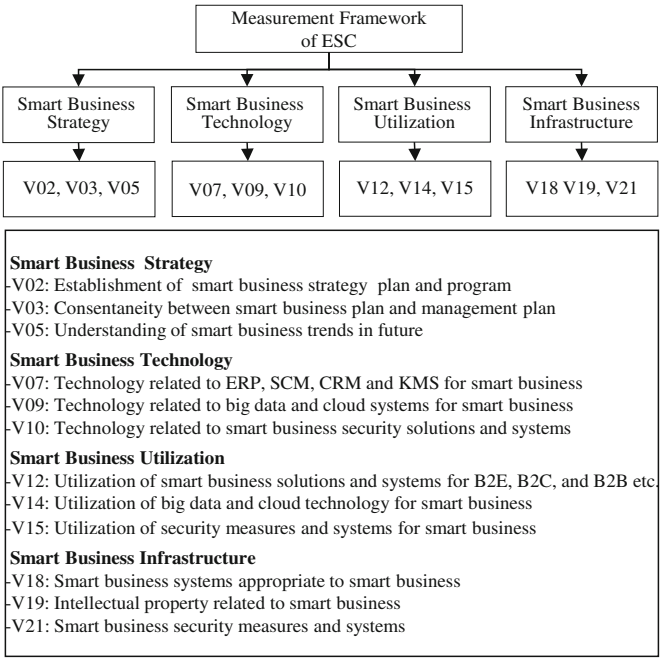


Fig. 1. The developed measurement framework for ESC.

measure ESC in terms of an entire smart business capability from smart business strategy to smart business infrastructure, including four measurement factors and twelve items.

Measuring ESC is a critical method to examine the entire smart business ability of a firm, based on its smart business strategy, technology, utilization, and infrastructure. We can use the framework to measure ESC across different industrial fields and business departments, and perhaps even as a global measure.

5 Conclusions

This study presents a comprehensive framework that can efficiently measure ESC to efficiently perform a firm's management activities and improve their performances and competitiveness in a global management environment. This research provided a reasonable framework that can measure perceived ESC from an entire smart business perspective. The developed framework with adequate validity and reliability presents groundwork for the development of a standard framework to measure for ESC.

Acknowledgments. This research was supported by a grant from the Academic Research Program of Korea National University of Transportation in 2016.

References

1. Hilty, L.M., Aebischer, B., Rizzoli, A.: Modeling and evaluating the sustainability of smart solutions. Environ. Model Softw. **56**, 1–5 (2014)
2. Heck, E.V., Vervest, P.: Smart business networks: concepts and empirical evidence. Decis. Support Syst. **47**, 275–276 (2009)
3. Chang, Y.F., Chen, C.S., Zhou, H.: Smart phone for mobile commerce. Comput. Stand. Interfaces **31**, 740–747 (2009)
4. Busquets, J., Rodon, J., Wareham, J.: Adaptability in smart business networks: an exploratory case in the insurance industry. Decis. Support Syst. **47**, 287–296 (2009)
5. Yoon, C.Y.: Development of a measurement model of personal information competency in information environment. Korea Soc. Inf. Process. Syst. **14-D**, 131–138 (2007)
6. Leem, C.S., Kim, S.K.: Introduction to an integrated methodology for development and implementation of enterprise information systems. J. Syst. Softw. **60**, 249–261 (2002)
7. Bharadwaj, A.S.: A resource-based perspective on information technology capability and firm performance: an empirical investigation. MIS Q. **24**, 169–196 (2000)
8. Jiao, H., Chang, C., Lu, Y.: The relationship on information technology capability and performance: an empirical research in the context of china's yangtze river delta region. In: Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management, pp. 872–876 (2008)
9. IBM Research Report: Business Services as a Modeling Approach for Smart Business Networks, 2 June 2006. www.ibm.com
10. IT Policy Research Series: Issue and Business Trend Change of Smart Era, vol. 3, pp. 1–13 (2011). www.nia.or.kr
11. Jang, H. K.: Change of Paradigm of Firm Business Environment provided by Smart Phone, DigiEco Focus (2010). www.digieco.co.kr
12. Lee, H.J.: A study on the business strategy of smart devices for multimedia contents. J. Inf. Process. Syst. **7**, 543–548 (2011)
13. Gabbanini, F., Burzagli, L., Emiliani, P.L.: An innovative framework to support multi-model interaction with smart environments. Expert Syst. Appl. **39**, 2239–2246 (2012)
14. Lee, S.H., Do, H.O., Seo, K.D.: A study on management plans for activating of smart work. Digital Policy Res. **9**, 245–252 (2011)
15. Lee, T.G.: Smart manufacturing execution system and business intelligence. J. Korea Inst. Inf. Technol. **9**, 35–43 (2010)
16. Doll, W.J., Torkzadeh, G.: The measurement of end-user's information satisfaction. MIS Q. **12**, 982–1003 (1988)
17. Torkzadeh, G., Doll, W.J.: The development of a tool for measuring the perceived impact of information technology on work. Omega, Int. J. Measur. Sci. **27**, 327–339 (1999)
18. Torkzadeh, G., Lee, J.W.: Measures of perceived end-user's information skills. Inf. Manag. **40**, 607–615 (2003)

Occluded Pedestrian Classification Using Gradient Patch and Convolutional Neural Networks

Sangyoon Kim[1,2(✉)] and Moonhyun Kim[2(✉)]

[1] Samsung Electronics, Suwon, Gyeonggi-do, South Korea
sangyun7.kim@samsung.com
[2] Sungkyunkwan University, Suwon, Gyeonggi-do, South Korea
mhkim@skku.edu

Abstract. Occlusion handling has been an important topic in pedestrian recognition. This paper proposed new approach for occlusion handling by Gradient Patch and Convolutional Neural Network (CNN). There are several researches of occlusion handling use parts annotations or manual labeling of body parts. However our method is learning partial features without any prior knowledge. Our model is trained parts detector with multiple of partial features that selected by gradient patch. Gradient patch compute the orientation of the edge in sub-region and find the extra partial features along the edge directions. Our experiments represented the effectiveness of Gradient Patch for occlusion handling in the INRIA and Daimler pedestrian dataset.

Keywords: Pedestrian classification · Partial detection · Gradient patch · Convolutional Neural Networks

1 Introduction

The pedestrian detection from single images has been studied in the past decade. Since this problem is important for various applications, such as intelligent vehicles, robotics and video surveillance.

In recent years, many classification approaches have been used for achieving the progress on pedestrian detection. VJ [17] and HOG [18] features methods are widely used with various cascade of boosted classifiers [11], SVM [14] classifiers. In recent researches the feature extraction part where include above handcraft features methods are changed to deep learning methods that unsupervised feature learning and learnable features [3, 8]. Most of all Convolutional Neural Network method [4, 16] achieved brilliant results in pedestrian detection [1, 2, 5, 6, 14].

However occluded pedestrian detection is still the obstacles. Occlusion handling is the key issue on pedestrian detection problem, because about 70 % pedestrian are occluded in one video frame, refer to the recent survey [10]. Currently Two representative approaches for occlusion handling are (a) learning each parts for different occlusion types [7, 10, 15]. (b) modeling part visibility as latent variables [12, 13]. First approach is training each part detector using annotations or statistics information of the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_32

occlusion patterns. Second one is divided pedestrian region into several parts and inferred each part's visibility using latent variables. While these approaches have achieved good results, learning model using prior knowledge and manually divided parts could be specify training dataset. It may not have stable performance on other types of occlusion dataset.

Our study aims to occlusion handling with simple partial features gain from gradient patch. The main advantage of our method does not require any prior knowledge like parts annotation or manual labeling of body parts. Also our model generate pedestrian classifier for occlusion handling without any artificially generated occlusion data [9]. As shown in Fig. 1, our system consists of two classifiers that generate by full patch and gradient patch. The holistic classifier trained with whole pedestrian image to reduce candidate windows. And the partial classifier trained partial features to recognize occluded pedestrians. We evaluate our proposed method on two pedestrian dataset (INRIA and Daimler) and show the effectiveness of Gradient Patch.

2 Proposed Method

The proposed method for learning pedestrian classifier with gradient patch is shown in Fig. 1. We use two steps to classify pedestrians like cascade classifier [11]. First we use holistic classifier and then if the result is not confidence then use partial features classifier to recognize occluded region. We train a holistic classifier with non-occluded pedestrian dataset. And we train a partial classifier with sub-region that we selected by gradient patch for partial features classifier.

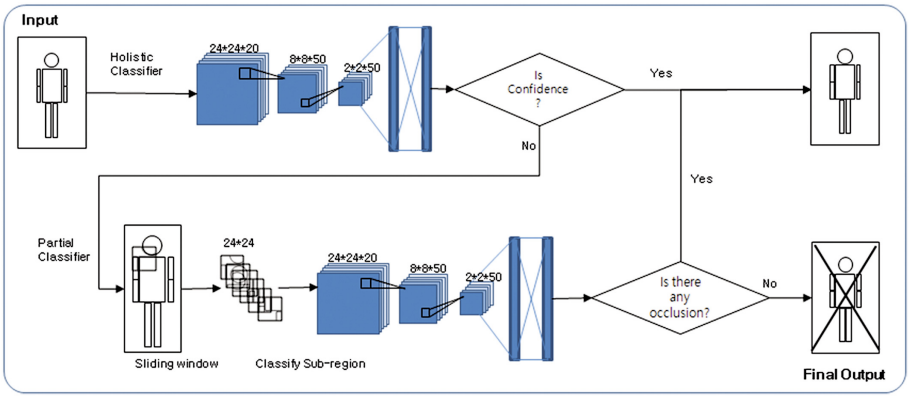


Fig. 1. A proposed method for classify pedestrians. Top model is holistic classifier for reduce candidate windows and bottom model is partial classifier for occlusion handling.

2.1 Gradient Patch for Partial Features

As shown in Fig. 2, Gradient patch calculate edge orientation of sub-region to find the important parts for partial features. In the begin of research, we consider using

randomly selective patch, however the false positive rate was too high, due to useless partial features, such as center of body and background. To handle this problem, we use a gradient condition to select patch location. Compute the gradient of sub region and then select extra patch along the orientation of edges. We use HOG [18] features with 6 orientation bins and 28×28 blocks and cells. For the fine condition test, we manually choose the initial location as left, top and bottom of body to easily find the extra gradient patch.

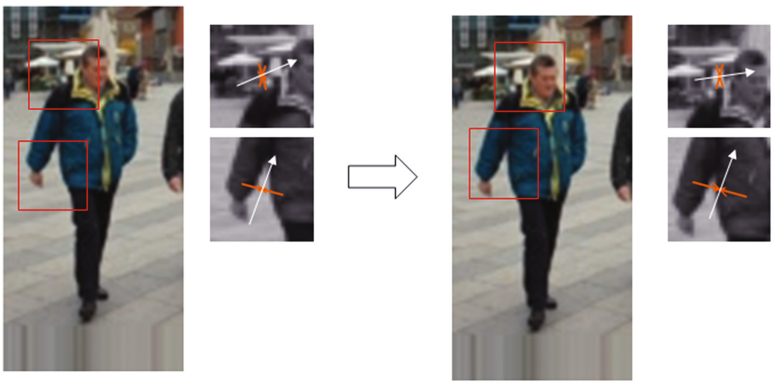


Fig. 2. Select partial region by Gradient Patch on INRIA dataset (Orange lines are result of HOG visualization and white lines are actual move direction)

2.2 Design Convolutional Neural Network

We use Convolutional Neural Network to training full features and partial features. CNN [6] has been successfully applied in pedestrian detection. The architecture of our model is presented in Table 1. We use small network to training full features and partial features classifier. We modified CifarNet parameters to change filters number, kernel size and remove last pooling layer to fit our patch size (28 × 28). CifarNet [14] is a small network that performing 2 % better than the AlexNet on Caltech dataset which has 600 times more parameters than CifarNet.

Table 1. Parameters of Convolutional Neural Network

Type	Filter/Stride	Filters number	Output map size 28 × 28
Conv1	5 × 5/1	20	24 × 24
Pool1	2 × 2/2	–	12 × 12
Conv2	5 × 5/1	50	8 × 8
Pool2	2 × 2/2	–	4 × 4
Conv3	3 × 3/1	80	2 × 2

3 Experiment Results

Our purpose is not only learning pedestrian model with occlusion handling, but also measuring the performance on non-occluded pedestrians. We choose two famous public pedestrian datasets (INRIA and Daimler). We assume that pedestrian candidate windows already given, since we want to focus classify performance using partial features that selected by gradient patch.

The INRIA and Daimler dataset consist of 2,416 pedestrians and 12,180 non-pedestrians, 52,112 pedestrians and 32,465 non-pedestrians training samples respectively. Also the Daimler dataset contains 11,160 partially occluded pedestrians separately.

3.1 Random Patch Vs Gradient Patch

In this work, we evaluate the performance of the gradient patch and random patch method on INRIA dataset. Also we compare to different number of patch. Figure 3 report on the performance of each types and patch number. The results show that proposed method is the best performance on non-occluded dataset and gradient patch method is better than random patch. In order to know the effectiveness of patch number, we train more set of gradient patch from 6 to 12, but the result show that the performance is not always improve. Because the human shape has many of vertical features, so the gradient patch usually find vertical features.

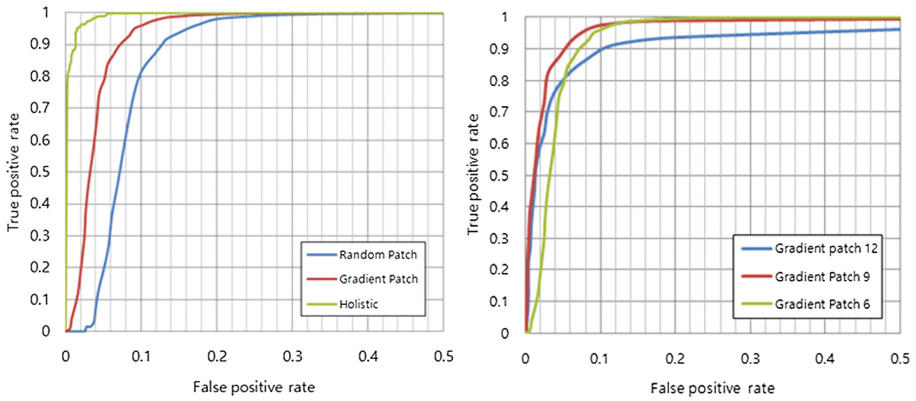


Fig. 3. Comparison of patch types and patch number. Label number indicates the number of patch.

3.2 Evaluation on Partially Occluded Dataset

We evaluate the performance of occlusion handling on Daimler dataset. The proposed method compared with Gradient Patch classifier and full features classifier. Results on terms of ROC performance are given in Fig. 4. The best classifier is proposed method

which is combined with gradient patch and holistic model. The result shows that the Gradient Patch improved 8.7 % true positive rates. In Fig. 5 shows the result of occlusion handling.

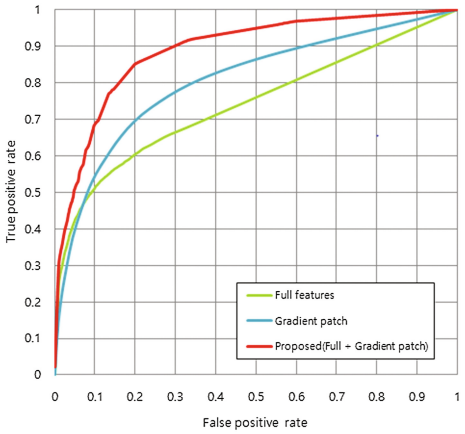


Fig. 4. Comparison of proposed method to gradient patch on Daimler dataset (partially occluded dataset).

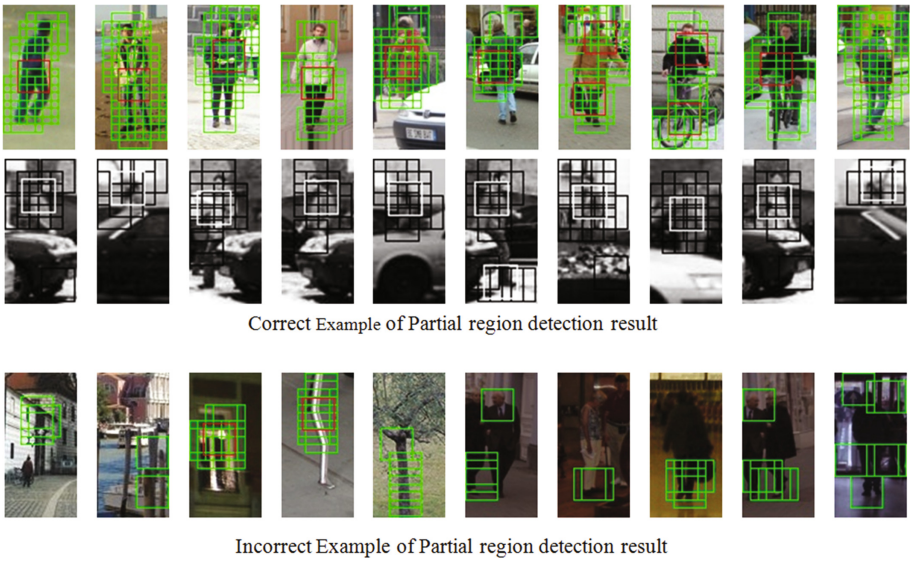


Fig. 5. Example of classification results of our approach on public dataset (e.g. INRIA and Daimler). Correct detection partial feature are green and merged 4 box in red.

4 Conclusion

In this paper, we proposed Gradient Patch to improve the performance of pedestrian classification with occlusion handling. Using holistic classifier in the first stage method highly improve the performance, due to reduce false positive result. Future work, we expect more improvement with different patch size and form, also we apply pre-trained model and fine tuning method on CNN to performance.

References

1. Ouyang, W., Wang, X.: Joint deep learning for pedestrian detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 2056–2063 (2013)
2. Sermanet, P., Kavukcuoglu, K., Chintala, S., LeCun, Y.: Pedestrian detection with unsupervised multi-stage feature learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3626–3633 (2013)
3. Bengio, Y.: Learning deep architectures for AI. Found. Trends Mach. Learn. **2**, 1–127 (2009)
4. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Neural Information Processing Systems (2012)
5. Chen, X., Wei, P., Ke, W., Ye, Q., Jiao, J.: Pedestrian detection with deep convolutional neural network. In: Jawahar, C.V., Shan, S. (eds.) ACCV 2014. LNCS, vol. 9008, pp. 354–365. Springer, Heidelberg (2015). doi:10.1007/978-3-319-16628-5_26
6. Dollar, P., et al.: Pedestrian detection: an evaluation of the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **34**(4), 743–761 (2012)
7. Mathias, M., Benenson, R., Timofte, R., Van Gool, L.: Handling occlusions with franken-classifiers. In: ICCV (2013)
8. Bengio, Y., Courville, A., Vincent, P.: Representation learning: a review and newperspectives. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798–1828 (2013)
9. Aly, S., et al.: Partially occluded pedestrian classification using part-based classifiers and Restricted Boltzmann Machine model. In: 2013 16th International IEEE Conference on Intelligent Transportation Systems-(ITSC). IEEE (2013)
10. Tian, Y., et al.: Deep learning strong parts for pedestrian detection. In: Proceedings of the IEEE International Conference on Computer Vision (2015)
11. Cheng, W.-C., Jhan, D.-M.: A cascade classifier using Adaboost algorithm and support vector machine for pedestrian detection. In: 2011 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE (2011)
12. Ouyang, W., Wang, X.: Joint deep learning for pedestrian detection. In: Proceedings of the IEEE International Conference on Computer Vision (2013)
13. Ouyang, W., Wang, X.: A discriminative deep model for pedestrian detection with occlusion handling. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE (2012)
14. Hosang, J., et al.: Taking a deeper look at pedestrians. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
15. Enzweiler, M., et al.: Multi-cue pedestrian classification with partial occlusion handling. In: 2010 IEEE Conference on Computer vision and pattern recognition (CVPR). IEEE (2010)

16. Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y.: Over Feat: integrated recognition, localization and detection using convolutional networks. In: International Conference on Learning Representations (2014)
17. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**, 137–154 (2004)
18. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893 (2005)

A Design of Secure Authentication Method with Bio-Information in the Car Sharing Environment

Sang-Hyeon Park, Jeong-Ho Kim, and Moon-Seog Jun(✉)

Soongsil University, Seoul, Republic of Korea
{shyeonl5,kimpocjstkl,mjun}@ssu.ac.kr

Abstract. Car sharing services have become a new form of public transport after the financial crisis as consumers' perceptions changed and their consciousness about environmental preservation and smart phone penetration increased. Countries overseas have already seen these services catch on with many users. In Korea, too, a pilot operation started in 2013 and now car sharing services boast about 200,000 members. Although the market evolved with many users, the security part is lagging behind. To rent a car, you only need to log in and you can use a smart key to open the car door and drive it. Since the simple ID/PW authentication method has many issues, a stronger authentication is needed to offer reliable services.

Keywords: Car-sharing · Authentication · Bio-authentication

1 Introduction

Car sharing services are a type of convergence services between the automobiles and IT industries. By combining the automobile with wireless communication and payment services, no driver is needed for services offered 24/7. Whenever a member wants a car, they can reserve it by time segments [1, 2]. It is an ultra short-term can rental business and is now emerging as a new form of public transportation. In Korea, it was adopted in 2013 and currently has over 200,000 members. The market is growing as consumers become more environment-conscious and more practical in their consumption patterns, seek to save costs and prefer convenient services using their smartphones which have seen a fast increase in penetration. The market is particularly growing in North America and Europe. The US has over a million members in car sharing services, while in Europe it is anticipated that the number of members will grow from the current 500,000 to over 2.4 million within three years. The Korean market, compared to its European or American counterpart, had a late start but it has the highest usage per capita. Despite these market trends, there are weaknesses. The largest service provider in Korea uses an ID/PW method when reserving or using cars, which is vulnerable to security. Since users prefer ID/PWs that are familiar to them, a stranger can guess them. In most cases, continued password management is not undertaken. Even if the user is cautious, malicious attackers can extract the password through farming, snipping or phishing. If a user's account is stolen, then the malicious attacker can use the victim's account for car

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_33

services. This can lead to small monetary damage caused by the user's credit card being charged to larger damage if there is an accident. There is also a loophole for people who cannot drive to rent a car, leading to damage to infrastructure and physical damage to innocent victims. Since car sharing services are fast becoming a mainstream public transport mode, it is clear that there is a growing market for environment-friendly and practical services. The security aspects of the services, however, must also be improved as the market grows. This paper seeks to suggest a more powerful authorization method using bio-information.

2 Relation Study

2.1 Tele-Authentication

Tele-authentication was suggested by Lamport in 1981 [3]. Lamport proposed a method where the user is authorized in an channel where the remote server is open. Based on Lamport's studies, many studies on password authentication protocol that is low-cost and highly efficient are being conducted [4–7]. The security requirements of the password authentication protocol fall into six categories.

1. Security against Password-Guessing Attacks
   A password estimation attach is a preemptive attack where a dictionary file is made of passwords commonly used by users is applied.
2. Security against Replay Attacks
   Replay attacks save the messages that a legitimate user had used in previous sessions and re-transmit them to gain authentication from the server.
3. Security against Impersonation Attacks
   Impersonation attacks occur when the attacker takes part in the communication session to disguise himself as a legitimate user and gain authentication from the server.
4. Security against Stolen-Verifier Attacks
   The attacker who has stolen a password verifier disguises himself as a legitimate user to gain authentication from the server.
5. Security against Denial of Service Attacks
   Denial of service attacks are undertaken by large quantities of traffic to the server to disrupt the legitimate users.
6. Provision of mutual authentication
   Mutual authentication verifies the other party's identity from both sides of the user and the server.

2.2 Bio-Authentication

Bio authentication technology was developed as a powerful identification tool since 9/11 terrorist attacks. In accordance with international standards for electronic passports, it has now become a global tool used in internationally recognized ID cards [8]. With this development, smartphone-based mobile bio identification technology also

evolved [9, 10]. Finger print recognition sensor was installed onto smartphones, which led to bio authentication technologies become integrated into our daily life. These technologies are applied to various services including mobile payment and unlocking of applications.

3 Propose

The suggested protocol will use finger print recognition for authentication. Table 1 shows the abbreviations to be used in the protocol.

Table 1. Abbreviation in protocol

Abbreviations	Meaning
B_U	Use's fingerprint information
ID_U	Use's ID
$E_{Xpublic}$	Encrypt X's public key
$D_{Xprivate}$	Encrypt X's private key
D_{SN}	Device serial number
R_X	X's random number
U_N	User's name
K	User&Server symmetric key

The suggested protocol is as follows. It is assumed that the user has first joined as a member on the server and the server knows the user's ID and Password.

3.1 Registration Protocol

Step1. The user's fingerprint information is acquired from the device and stored in the SE.

Step2. IDU and BU are linked to sash. DSN and RU are encrypted using the open key, and then transmitted to the server.

Step3. The server uses the personal key to decrypt the encrypted message sent by the user and items are acquired.

Step4. In order for the server to match the user with the device, the communication services provider received the user's name and the serial number of the device.

Step5. Communication Services provider check user's name and the serial number of the device.

Step6. If the user and the device serial number received match, the service provider sends "True", and otherwise it sends a "False" value.

Step7. If the value sent from the service provider is True, then h(IDU||BU) and DSN values are linked and hashed to generate a symmetric key, K. If the value is False, then an error message is sent.

Step8. The server, too, in order to verify the user, encrypts using the open key the random number sent b the user and the random number that it generated, and transmits them.

Step9. Using it's own individual key, decryption is carried out and its own random number and the server's random number are received. Then it is verified whether the random number it sent is correct. If it is, then h(IDU‖BU) and DSN values are linked and hashed to generate the symmetric key, K.

Step10. Using the server's open key, the random number of the server is sent for verification and the registration process is complete (Fig. 1).

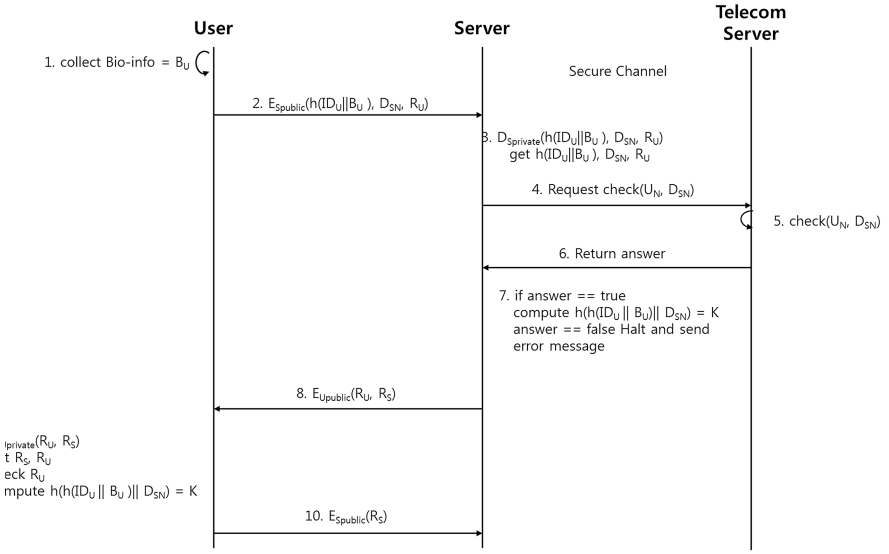


Fig. 1. Registration protocol

3.2 Login, Smart Key Authentication Protocol

When logging in for reservation or for using the smart key, it is only possible through the registered smartphone device. The symmetric key generated during registration is used for authentication of service usage.

Step1. The fingerprint sensor is used to collect fingerprint information. Using the collected information and the device's own serial number, a symmetric key is generated.

Step2. The symmetric key is used to encrypt the ID, PW, and random value which are transmitted.

Step3. The symmetric key is used for decryption. It is verified whether the ID and PW are correct, and the server's random value is generated.

Step4. The random value received from the user and the random value generated by the server are sent using the symmetric key.
Step5. The symmetric key is used to encrypt the received message and to check the random value.
Step6. The server's random value is encrypted using the symmetric key and the authentication is complete (Fig. 2).

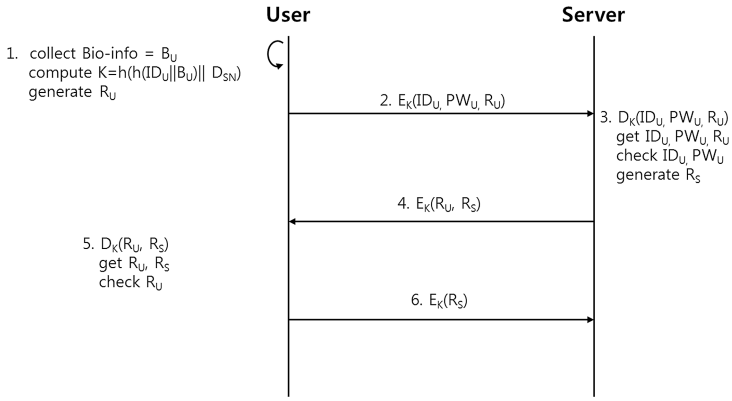


Fig. 2. Login, smartkey authentication protocol

4 Security and Performance Analysis

The suggested protocol provides services through fingerprint recognition technology, in order to overcome the vulnerabilities of the simple ID/PW authentication system so that more secure services can be offered.

1. Replay attacks
   To prevent replay attacks, the suggested protocol generates a random value for each registration session.
2. Counterfeit attacks
   The suggested protocol, in addition to using open key-based passwords, uses the challenge-response method to send again the value that the user had transmitted for verification. The value sent by the server, too, is re-transmitted again to make it fool-proof.
3. Intermediary attacks
   Although the intermediary can dupe the server and use its own open key to generate the key of each between the user and the server, but by connecting with the service provider to match the user information with the device information, more resistance is built against intermediary attacks.
4. Farming, snipping
   Even though the user's ID and PW are stolen through farming or snipping, the fingerprint information cannot be collected and therefore illegitimate users cannot be authorized.

5 Conclusion

Car sharing services offer a type of automobile rental services where the user can rent a car by the minute or the hour. The existing car sharing services simply use a login system to reserve a car and a smartkey is used for the user to board and drive the car. When there is an attack via farming or snipping, the user's ID and Password can be stolen, and the attacker can rent and even drive the car. The victim user would have to not only pay for the rental fees, but also bears the responsibility when there is an accident. If an attacker who is not qualified to drive steals an account, it can lead to accidents with casualties. If the authentication method as suggested in this paper is used, the fingerprint information can offer a more powerful authentication method and prevent illegitimate usage of services.

References

1. Adam, M.-B.: Car-Sharing: where and how it succeeds, vol. 108. Transportation Research Board (2005)
2. Meijkamp, R.: Changing consumer behaviour through eco-efficient services: an empirical study of car sharing in the Netherlands. Business Strategy Environ. **7**(4), 234–244 (1998)
3. Lamport, L.: Password authentication with insecure communication. Commun. ACM **24**(11), 770–772 (1981)
4. Chang, C.-C., Wu, T.-C.: Remote password authentication with smart cards. IEEE Proc. Comput. Digital Tech. **138**(3), 165–168 (1991)
5. Hwang, M.-S.: A remote password authentication scheme based on the digital signature method. Intl. J. Comput. Math. **70**(4), 657–666 (1999)
6. Hwang, M.-S., Lee, C.-C., Tang, Y.-L.: An improvement of SPLICE/AS in WIDE against guessing attack. Informatica **12**(2), 297–302 (2001)
7. Wang, G., Bao, F.: Cryptanalysis of timestamp-based password authentication schemes using smart cards. In: Ning, P., Qing, S., Li, N. (eds.) ICICS 2006. LNCS, vol. 4307, pp. 399–409. Springer, Heidelberg (2006). doi:10.1007/11935308_28
8. Kim, et al. Tele biometrics-based authentication technologies non-face-to-face Standardization. Inf. Secur. J. **25**(4), 43–50 (2015)
9. Cassone, J.: Personal biometric authentication and authorization device. U.S. Patent No. 6,983,882. 10 Jan 2006
10. Liu, S., Mark, S.: A practical guide to biometric security technology. IT Professional **3**(1), 27–32 (2001)

A Design of Certificateless-Based Device Authentication Scheme in the SmartHome Environment

Jae Seung Lee[1], Jaehwa Chung[2], and Sangkee Suk[3(✉)]

[1] Soongsil University, Seoul, Republic of Korea
ljs0322@ssu.ac.kr
[2] Korea National Open University, Seoul, Republic of Korea
jaehwachung@knou.ac.kr
[3] Seoul National University of Science and Technology, Seoul, Republic of Korea
sksuk@seoultech.ac.kr

Abstract. Recently, the wireless communication technology and sensor devices are utilized in many fields with smart-home market growth. The IoT environment collects various and vast range of device information for intelligent service, provides service based on user information, controls devices, and utilizes dissimilar devices. However, along with the development, the security threat in smart-home environment occurs frequently. In reality, the proof point and HP published the seriousness of weak security and damage cases in smart-home environment. Therefore, this study proposes smart node and certification method based on Certificateless signcryption between smart devices for remote control to solve the security problem occurring in smart-home environment.

Keywords: IoT · HomeNetwork · SmartHome · Device authentication

1 Introduction

Recently, the number of smart devices is increasing geometrically due to the development of communication technology and smart device, which led to expansion of service and development area of smart device, and resulted in rapid growth of global smart-home market. According to the actual market research institute, the Strategy Analytic, the smart-home market scale of USA in 2012 is supposed to be 7.6 billion dollar which is year on year 55.1 % increase, and it is expected to be 115 billion dollar by 2019 with 19 % of yearly growth. Also, the European smart-home market scale is supposed to be 3.1 billion dollars which is 82.4 % year on year growth, and it is expected to be 10.3 billion dollars by 2016 with 26.9 % yearly growth. The smart-home market growth is spreading with various business development, and in the situation where various fields of corporate such as home appliances, communication and security service, mobile, and utility are actively participating in smart-home market, IoT with rapid growth, M2M and various sensors, wearable computing, movement and voice recognition technology are fused, and the complex device environment is constructed to enable the proposal of various residential environmental service.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_34

Likewise, the smart-home market has grown and utilized in various fields, but the security threat is also becoming an issue. Current smart-home service concept is no better than “home networking” or “connected home” but, by the introduction of app store or market of smart device, it is expected to pass the bounds of one-direction service structure focusing on the supplier, in which case when the ecosystem is adapted to smart-home, the security of smart device must be essentially considered.

Actually, the Silicon Valley security service business, the Proof Point announced that the 750 thousand cases of phishing and spam emails sent last years was through smart-home IoT product hacking. The research report published by HP on February 2015 warned that “most of the smart-home IoT equipment has weak security in encryption and certification process”, and “because one must provide with personal information to use smart-home IT equipment, it became more vulnerable to cyber crime”.

Therefore, this study proposes smart node and certification method based on Certificateless signcryption between smart devices for remote control to solve the security problem occurring in smart-home environment. The proposed certification method prevents smart node access of malicious user by mutual certification, and also increases the device energy efficiency with Certificateless signcryption which is lighter than SSL communication which is difficult for application.

2 Related Work

2.1 Certificateless Signcryption

In 2008, Barbosa and Farshim first proposed the CLSC method [1] which is certificateless, and later Wu and Chen introduced efficient CLSC scheme [2], but Selvi later verified that it was insecure. Xie and others have proposed CLSC scheme using 2 pairing calculation on signcrypt and unsigncrypt level by bilinear map [3], but this scheme is rather difficult to apply to actual filed due to its high pairing calculation with calculation cost. Therefore, Xie and Zhang proposed Certificateless signcryption. This study uses Certificateless signcryption method without pairing proposed by Xie and Zhang [4]. The Certificateless signcryption method is as below.

- Setup: Create params, the system parameter and Master-Key from random k.
- Partial-Key-Extract: Create partial individual key D_{ID} and partial public key P_{ID} from parmas obtained by algorithm, Master-Key and user recognition value ID.
- Set-Secret-Value: Create secret value S_{ID} by params and user recognition value ID.
- Set-Public-Key: Create user public key PK_{ID} by params, user ID_A and secret value $S_{ID.}$
- Set-Private-Key: User individual key SK_{ID} is created from params, user partial individual key D_{ID} and secret value S_{ID}.
- Encrypt: Create coded message σ from params, individual key sk_ID_s of receiver and message m, which is σ = signcrypt(params, $SK_{ID,}$ ID_R, PK_{ID}, m).
- Decrypt: Message m is restored from params, recognition value IDs of transmitter, public key PK_{ID}, individual key SK_{ID} of receiver and code message σ, and when the code message σ is verified to be correct, the message m is restored and if not, the decrpyting fails. In other words, when = unsigncrypt(params, ID_S, PK_{ID}, SK_{ID}, σ,

and P is decrypted in normal condition, message m is output or else, the error is output.

- The algorithm of Setup and Partial-Key-Extract is conducted by KGC. The partial individual key D_{ID} and partial public key P_{ID} is transmitted to user through secret channel, and the algorithm creating user's public key and individual key and Set-Secret-Value algorithm are conducted by user.

3 Proposed Scheme

(3−1) is a device certification method based on Certificateless signcryption public key in IoT home network environment. The characteristic of Certificateless signcryption public key is that it uses device ID to create public key. At this point, if user provides distinct information value of device, it is compared with the ID value sent from server and verifies the link to public key. Then, the ID of Device that needs certification is received through Home Network, the public key and ID received by server is verified KGC to access to correct service, and therefore verify that the public key was received accordingly. Unlike the Certificateless signcryption public key method on equipment, when utilizing Certificateless signcryption public key method to general smart-home service, KGC takes charge of public key certification additionally.

3.1 Certification and Verification

- First, after security parameter k is entered, the KGC creates p and q. At this point, p and q must satisfy q > 2 k and q|(p−1). Then, after selecting g, the Master Key x is randomly selected to calculate $y=g^x$. Next, the hash function H1, H2, H3, H4, H5 are calculated and reveal system parameter params = <p,q,n,g,y, H1, H2, H3, H4, H5).
- The service provider obtains parameter, master x, and user recognition value ID∈ {0,1}*. Later, wID = grID, dID = rID + xH1(ID,wID), vID = gr′ID, σID = r′ID + xH2(ID, wID, vID) is calculated and return the partial secret key DID and PID = (wID, vID, σID).
- User params value and ID is entered, the KGC selects z ∈R Z*q and creates $s_{ID} = z$.
- Service provider uses partial public key p_{ID} and secret value s_{ID} to create user individual key $sk_{ID} = (d_{ID}, s_{ID})$.
- To create public key responding to individual key, the service provider enter params, user partial public key P_{ID} and secret value S_{ID} to calculate $\mu_{ID} = g^{sID}$ and create public key $Pk_{ID} = (\mu_{ID}, w_{ID}, v_{ID}, \sigma_{ID})$.
- To send message to service provider with recognition value of receiver and value key, the certification process described below is conducted.
- First, verify the $g^{\sigma R} = v^B y^{H2(B,WB,vB)}$. Later, random number r is created to calculate $h = H_3(m, t, \mu^rB, w^rB, y^{H1(B,wB}r, \mu_A, w_A, \mu_B, w_B)$, $h' = H_4(m, t, \mu^rB, w^rBy^{H1(B,wB)r}, \mu_A, w_A, \mu_B, w_B)$, $s = r\text{-}hd_A - h's_A$ and send coded message $\sigma = (c, s, t)$.

- The service provider verifies $g^{\sigma A} = \nu_A y^{H2(A,\omega A,\nu A)}$. First, $m = c \oplus H_5(t^{sB}, t^{dB})$ is calculated. Then when $h = H_3(m, t, t^{sB}, t^{dB}, \mu_A, \omega_A, \mu_B, w_B)$ and $h' = H_4(m, t, t^{sB}, t^{dB}, \mu_A, w_A, \mu_B, w_B)$ is satisfied, the certification is complete.

4 Security Analysis

In case of confidentiality and integrity, the individual key obtained by mutual certification of device and public key. When the replay attack or message tampering attack occurs, the newly created session key rather than the session key at the time of message tempering and time stamp provides security from replay attack. When sniffing, the messages sent to each node use session key between the renewed node to apply encryption and sent which makes it secure. In case of spoofing, the nodes are already mutually certified, so that when spoofing attack occurs, the secret key between nodes is unavailable and secure.

5 Conclusion

The development of communication technology and smart device led to the increase of smart-home service market and commercialized various smart-home services. However, the security application condition is still incomplete and the vulnerability and damage cases have been reported to research institutes. Therefore, this study proposes protocols enable to respond to various security threat by the smart node and certification method based on Certificateless signcryption between smart devices for remote control in smart-home environment. Also, the non-pairing calculation method provided high-level security with better energy efficiency that the previous SSL communication.

References

1. Barbosa, M., Farshim, P.: Certificateless signcryption. Cryptology eprint archive:report 2008/143. http://eprint.iacr.org/2008/143
2. Wu, C., Chen, Z.: A new efficient certificateless signcryption scheme. In: International Symposium on Information Science and Engieering, ISISE 2008, vol. 1, pp. 661–64 (2008)
3. Xie, W., Zhang, Z.: Efficient and provably-secure certificateless signcryption from bilinear maps. Cryptology ePrint Archive: Report 2009/578. http://eprint.iacr.org/2009/578
4. Xie, W., Zhang, Z.: Certificateless signcryption without pairing. Cryptology ePrint Archive: Report 2010/187. http://eprint.iacr.org/2010/187

A Design of Secure Authentication Method Using Zero Knowledge Proof in Smart-Home Environment

Geunil Park, Bumryoung Kim, and Moon-seog Jun(✉)

Soongsil University, Seoul, Republic of Korea
{higa_psl5, gflawer, mjun}@ssu.ac.kr

Abstract. As the IoT technology is being changed, diverse services are being developed. The smart home environment, which is the representative IoT technology, is facilitating the life more convenient and the smart home market is emerged rapidly while the IoT technologies are applied, but when authenticating in smart home network, the authentication should be made using the secret key of the node. In the smart home network environment, the more the nodes are increased, the more the burden of home gateway to manage the secret key of the nodes is increased. To solve this problem, this study suggests secure and efficient authentication technique through the zero-knowledge proof without using the secret key between the node and the home gateway.

Keywords: IoT · IoE · Smart home · Authentication · Zero knowledge proof

1 Introduction

Recently, as the IoT technology is being developed and popularized rapidly, diverse and convenient services are developed by applying IoT technologies. It is being applied to various domestic and overseas industrial areas, and not only the global enterprises but also the startup enterprises are investing in it actively [1]. It refers to the environment that provides the human-oriented services by forming a network of the home appliances such as T.V., refrigerator, boiler, etc. in the smart home environment, which is the representative environment that provides convenient services that the diverse objects are communicating each other without human intervention or that the user can control the home appliances within home using the application [1, 2]. Through the smart home service, the user can be provided with the more secure, and convenient and pleasant entertainment, security, healthcare, energy services with low the maintenance cost than existing home environment. Generally, the home appliances in the smart home environment are referred to as IoT device, and when these IoT devices communicate with other IoT devices, they are authenticated through the wireless internet by making the secret key of the IoT device as the data. Since in the smart home network, IoT device is not a single node, when authenticating various IoT devices, Home Gateway must have the secret key of relevant IoT devices but there is problem that the home gateway should manage lots of secret keys. To solve this problem, this study suggests the more secure and efficient authentication through the zero-knowledge proof

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_35

suitable for smart home environment that the home gateway can do without having secret key of IoT device rather than authentication using private key of IoT device [3, 4]. In Sect. 2, the structure and existing smart home network, the problems and the vulnerability of the smart home network will be introduced, and in Sect. 3, secure and efficient authentication technique and the command protocol though the zero-knowledge proof between the node and the home network in the smart home network will be explained. In Sect. 4, the security assessment using suggested authentication method will be performed and in Sect. 5, the conclusion will be made.

2 Related Work

2.1 Smart Home Infrastructure

Smart home refers to the future residential environment that connect various information devices with the network and provides human-oriented services. Smart home network has a structure that authenticates and manages the IoT devices within home connected to home gateway and the user accesses the home gateway controlling the IoT devices without being limited by time and place [5]. In this structure, the various nodes authenticate with the private key such as ID and Password to authenticate the home gateway and the home gateway manages the private key table for the various nodes [4]. This structure has a problem that the private key is exposed to the wireless environment. In addition, as the smart home service are gradually diversified and the number of devices for the nodes are increased proportional to the needs of user, of which problem is that since the more the private key of various nodes are increased, the bigger storage is required increasing the cost [3, 4] (Fig. 1).

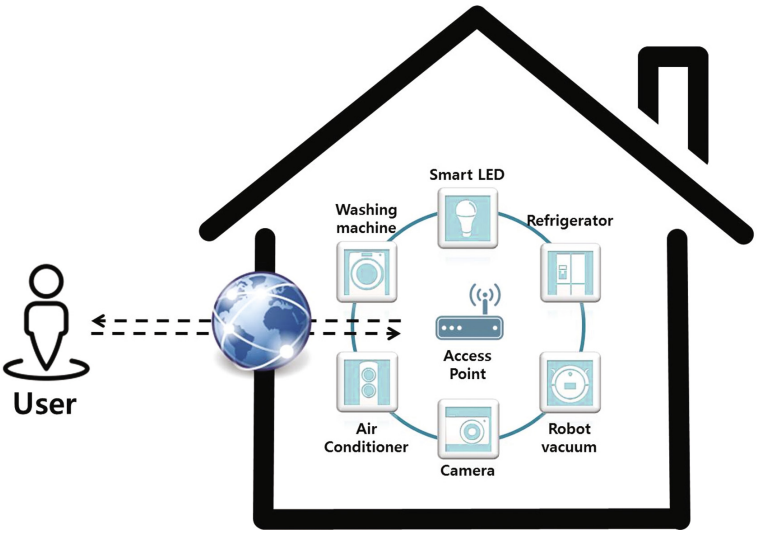


Fig. 1. Smart Home Infrastructure

3 Proposed Scheme

In this chapter, We show the proposed protocol which is secure and efficient between devices without any classified information.

To have effective certification, first distinguishing Node and exchanging key and after when it got order from registering of identification and registration protocol, it suggest command and authentication protocol which deliver command through zero-knowledge proof from Home Gateway and node. Moreover, this suggested protocol is called as what you can see from the Table 1.

Table 1. Scheme Notation

Criteria	Sub-criteria
ID_{node}	Node‘s Identifier
ID_{home}	Home Gateway‘s Identifier
n	Public Key for ZKP
Key_{gp}	Home Gateway Public Key
Key_{np}	Node Public Key
Key_{sy}	Symmetric Key
Key_{gs}	Home Gateway Secret Key
ML	Smart Home Management List
TS	Timestamp
$EKey_{us}$	User Secret Key
R	Random Number
Key_{ns}	Node Secret Key
Salt	Protect Random Number for Order

3.1 Smart Home Identification and Registration Protocol

The basic key exchange is made through the secure channel between the node and the gateway, and the node is identified and registered by user as shown in Fig. 2.

1. Node sends its ID to register it in the node list.
2. Home gateway combines its ID, open key and n value for open key, which is used for zero-knowledge proof, and sends to the node.
3. Node generates the symmetric key with the home gateway and combines with the open key of node and send back
4. Home gateway registers the identified node in the existing smart home’s node management list.
5. Home gateway proves the integrity from relevant home gateway by encrypting the renewed list with its secret key.

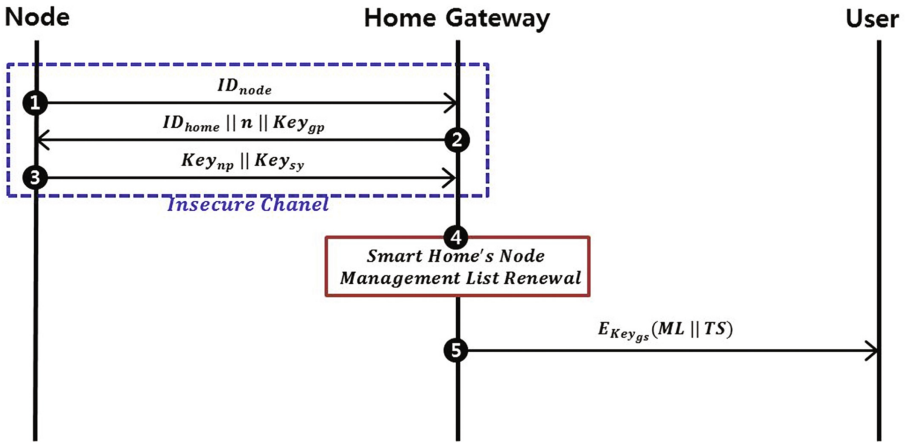


Fig. 2. Smart Home Identification and Registration Protocol

3.2 Smart Home Command and Authentication Protocol

Figure 3 shows the protocol to send the command to the registered node.

1. User encrypts the ID of node and the command with the user's secret key to send the command to the node, and encrypts them together with the Timestamp value with the open key of home gateway again and sends it.
2. Home gateway extracts the command and ID of node by decrypting the data received from the user with its secret key, Timestamp value and the user's open key and requests the authentication to relevant node.
3. Node selects a random number r of greater than 0 and less than n-1, and raises r to second power and put the module n to value x and send it. Here, n is the value of open key n used in the zero-knowledge proof.
4. As a step to send the challenge value in the zero-knowledge proof, in the home gateway, put random value of 0 or 1 to value c and send it.

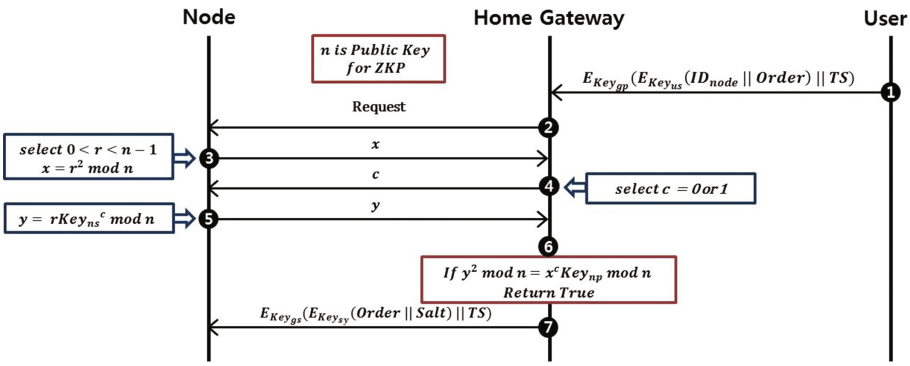


Fig. 3. Smart Home Command and Authentication Protocol

5. Raise value of r, which was selected in step 3, multiplied with the private key of node c^2 power, calculate the module and put it to value y, and send it to the home gateway.
6. Home gateway does not expose the secret key of node in the process of communication, but if the secret key held by node is normal, the below equation is valid, $y^2 mod\ n = xv^c mod\ n$ and the mode is authenticated.
7. It encrypts the order and salt values with the symmetric key exchanged before, and send it by encrypting it together with Timestamp value with the secret key of the home gate again. Here, since the command used in smart hone is not diverse and short, it provides the complexity by adding the random value to the salt value.

4 Security Analysis

4.1 Sniffing

Sniffing attack means what eavesdrops the packet exchange between other parties not its own in the network [6]. Simply speaking, it means the process of tapping the network traffic, and in the suggested protocol, since the secrecy of the data is guaranteed by encrypting or decrypting through the key initially exchanged, it provides the security against the sniffing attack.

4.2 MAC Address Spoofing Attack

To use the smart home services, the node should select AP (Access Point) and passes the authentication process of AP. Since the authentication process suggested by this study is to authenticate not with the secret information like MAC address [7] but through the zero-knowledge proof, it is secure from the MAC address spoofing attack, which releases the authentication by sending the forged message through spoofing MAC address.

4.3 Replay Attack

In case of replay attack, which the attacker disguises the rightful user using the valid data later after selecting and copying it in the protocol, since it sends each timestamp value together, the copied data cannot be used again after time elapsed [6]. In addition, since it encrypts the command adjacent to salt when delivering the command, it is secure against the replay attack.

5 Conclusion

Diverse and many services using IoT technology not only for the smart home services but also for the users are being increased. While the IoT devices are being increased, the gateway, which can manage and control lots of IoT devices, should have secret

information of many IoT devices, but to solve this problem, this study suggests the secure communication that reduces the burden of gateway by allowing the authentication using zero-knowledge proof without the secret information of the IoT devices or nodes.

References

1. Ye, X., Miao, X.: Smart Home IoT technology system based on ZigBee. Mod. Archit. Electric **9**, 009 (2010)
2. Reichherzer, T., Satterfield, S., Belitsos, J., Chudzynski, J., Watson, L.: An agent-based architecture for sensor data collection and reasoning in smart home environments for independent living. In: Khoury, R., Drummond, C. (eds.) Canadian AI 2016. LNCS, vol. 9673, pp. 15–20. Springer International Publishing, Switzerland (2016)
3. Zhang, W., et al.: Research status and development trend of smart grid. Power Syst. Technol. **13**, 004 (2009)
4. Valtchev, D., Frankov, I.: Service gateway architecture for a smart home. IEEE Commun. Mag. **40**(4), 126–132 (2002)
5. Byun, J., et al.: An intelligent self-adjusting sensor for smart home services based on ZigBee communications. IEEE Trans. Consum. Electron. **58**(3), 794–802 (2012)
6. Kahate, A.: Cryptography and Network Security. Tata McGraw-Hill Education (2013)
7. Stallings, W.: Network Security Essentials: Applications and Standards. Pearson Education India (2007)

Drone Classification by Available Control Distances

Mansik Kim[1], Hyungjoo Kim[1], Jungho Kang[1], and Jaesoo Kim[2(✉)]

[1] Soongsil University, Seoul, Republic of Korea
{mansik, hyungjoo.kim}@ssu.ac.kr, kjh7548@naver.com
[2] Seoul National University of Science and Technology, Seoul, Republic of Korea
jskim@seoultech.ac.kr

Abstract. A drone industry which started for military purposes is growing and are being spread in the private market as ICT technologies are developing. For example, it includes various industries such as leisure, construction, delivery, military, etc. and there are plenty of pioneer areas for future. However the researches for future drone developments have some challenges because there are no standards for drone developments, so the environment should be provided to develop and design them properly. This paper categorizes various purpose drones into 3 kinds of control distances by features, so it will help future drone development researches.

Keywords: Drone · Control distance · Aircraft · UAV · Classification

1 Introduction

The drone is an aircraft which can fly without pilot but by remote controller or itself [1]. It was usually developed for military purpose and now used in various private markets for enterprises or civilians. Because he or she is not on it, they are remotely controlled by wireless communication and the remote control method and mean vary according purposes and usage of drone. For example, most private drones fly in the area which the user can see it, the Amazon drone for delivery receives GPS signals and fly over kilometers to get specified location, and military drone can fly over hundreds or thousands kilometers using military satellite [2–4]. However there are no specific standards for these drones for a variety of uses, so the customized drone research is very difficult. This paper defined three kinds of remote control distances for drone to categorize them by features, so it can suggest some standards for drone to help future drone researches.

The paper consists as follows. Section 2 explains drone market trends categorized in three kinds of remote control distances and Sect. 3 presents a conclusion.

2 Drone Classification

This chapter shows drone classification into shot, middle, and long control distance.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_36

2.1 Short-Distance Control Drone

The distance range for short-distance control drone is from meter to tens meters which controller can see it in eyes. It usually uses short distance wireless communication such as Bluetooth, WiFi, etc. [5, 6] for remote control and does vertical take-off and landing. And also it has small body that contains small battery which makes it fly during only few or tens minute and computing power which can't send live video it take. So most short distance control drones are controlled depending on controller's eyes and the obstacle between them should not be existed. These are usually used in private market such as leisure activity, TV shooting from the sky, construction site, etc.

2.2 Middle-Distance Control Drone

The distance range for middle-distance control drone is from tens meters to kilometers which controller can't see it in eyes and dose vertical take-off and landing like short-distance control drone. It is controlled by AP or cellular system and is relatively expensive, because it can fly more and longer than short-distance control drone. Also it has bigger battery and computing power than short one so can deliver some objects and send live video to controller which replaces controller's eyes. Because the live video replaces controller's eyes, it usually has an obstacle avoidance technology or impact mitigation technique that prevents crash from other sight where camera can't see. They are used in the areas which human can't go in and Amazon is developing drone delivery service [3].

2.3 Long-Distance Control Drone

Long-distance control done can fly from kilometers to thousand kilometers even over nations. It uses wings to fly fast and far away than short and middle ones as well as it uses airstrip when it takes off and landing. They are usually designed with specific purpose before they are sold and use satellite for remote control. It use coal fuel so can make strong power and high computing resource which can do complex mission. Most of them are made for military purpose to patrol or attack specific target, but they are expected to be used in more various areas in the future because many regulation for them are being released and ICT technology are improving.

3 Conclusion

Drones were developed for military purpose and used in limited areas in the past but now it is used in various private industries because ICT technology is improved and regulation for them are released. However the drone researches for specific purpose are still very difficult because they are made without proper standards. This paper categorizes them into short, middle, and long distance control drone by features so it can help the drone researches for specific purpose.

References

1. Market Profile and Forecast: World Unmanned Aerial Vehicle Systems. Teal Group (2014)
2. Asaro, P.M.: The labor of surveillance and bureaucratized killing: new subjectivities of military drone operators. Soc. Semiot. **23**(2), 196–224 (2013)
3. Sheets, C.A.: China Beat Amazon Prime Air to the Commercial Drone Delivery Market. International Business Times (2013)
4. Evans, B.G.: Satellite Communication Systems. Institution of Engineering and Technology (2004)
5. Bluetooth SIG: Specification of the Bluetooth System, Core v1.1 (2001)
6. IEEE Computer Society LAN MAN Standards Committee: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications (1997)

A Design of Key Agreement Scheme Between Lightweight Devices in IoT Environment

Hague-Chung, Keun-Chang Choi, and Moon-Seog Jun(✉)

Soongsil University, Seoul, Republic of Korea
{standard, muziag, mjun}@ssu.ac.kr

Abstract. The IoT (Internet of Things) environment develops in which all the necessary information among things is exchanged due to the development of information and communications. The home IoT continuously develops, because of a merit that a user can control the home IoT remotely in the IoT environment. As the home IoT environment is built, communications using low specification devices, as well as high specifications devices, also increase. For safety communications in the home IoT environment, encryption algorithms, such as RSA providing message encryption and authorization, are required. However, they are difficult to be used for the low specification devices, where calculation function is limited in the home IoT environment. This study actually proposes the protocol by which low power and low specification devices communicate with user's smart devices through safe authorization procedure in the home IoT environment. The protocol proposed in this paper has a merit that it is safe, and it protects the re-use attack and middle attack.

Keywords: IoT · Smart home · Authentication

1 Introduction

The IoT environment recently develops exchanging all the information required among things through the Internet [1]. Concerning such an IoT environment, the home IoT environment products including switches, plugs, energy meters, thermostats, sensors detecting open and door locks are continuously released in the market. There is a convenience that the users in the home IoT environment can control devices in their homes with their own smart devices anywhere, anytime. A market research company, Strategy Analytics, predicted that the global smart home market would grow from USD 48 billion (KRW 54 trillion) in 2014 to USD 115 billion (KRW 129 trillion) in 2019 with annual average growth rate of 19 %. Korean market was also predicted to grow from KRW 6.8908 trillion in 2013 to KRW 18.2583 trillion in 2017 with annual average growth rate of 27.6 % [2]. Because ultra-small IoT devices are operated with low power in the home IoT environment, unlike existing PCs or mobile devices, computing performance is lower, and storage space is limited. IoT platform is mainly connected wireless, and thus, such protocols as ZigBee, Bluetooth, 6LoWPAN, Z-Wave, IEEE802.15.4, LoRA and Wi-Fi are used [3]. These protocols support diversely codes ranging from the codes with low security such as TDES and SHA-1 to the codes with high security such as RSA-2048 [4]. However, these standard protocols

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_37

do not support the code algorithms equipped with weight-light characteristics suitable for the IoT environment in reality. This paper proposes the protocol by which low power processors can communicate through safe authorization procedure. This paper is constructed as follows: Sect. 2 describes the relevant researches, and Sect. 3 introduces the system's protocol. Section 4 examines the implemented protocol's efficiency and security through the performance evaluation of the protocol. Lastly, Sect. 5 presents conclusion, and discusses improvements in the future protocol.

2 Relevant Researches

2.1 IoT

ITU presented the concept of IoT for the first time through the ITU-T World Summit on the Information Society in 2005. Existing information and communications technology enabled to send and receive information mutually between humans and things anywhere, anytime. A new concept, "anything," was added to the meaning of the current IoT as shown in Fig. 1. The meaning of IoT is currently defined as technology enabling connection and communications between humans and things, between humans and between things [5, 6]. Here, "anything" includes specific things in the physical space, authors or identified information in the virtual space. EU specified IoT as all things that can be connected and communicated with the surrounding environmental factors through intelligent interface having a unique identifier and virtual character [7]. In Korea in 2009, the Korea Communications Commission (Currently, the Ministry of Science, ICT and Future Planning) defined IoT as broadcasting and communications-fused ICT infrastructure, by which intelligent communications service between humans and things and between things can be used safely and conveniently in

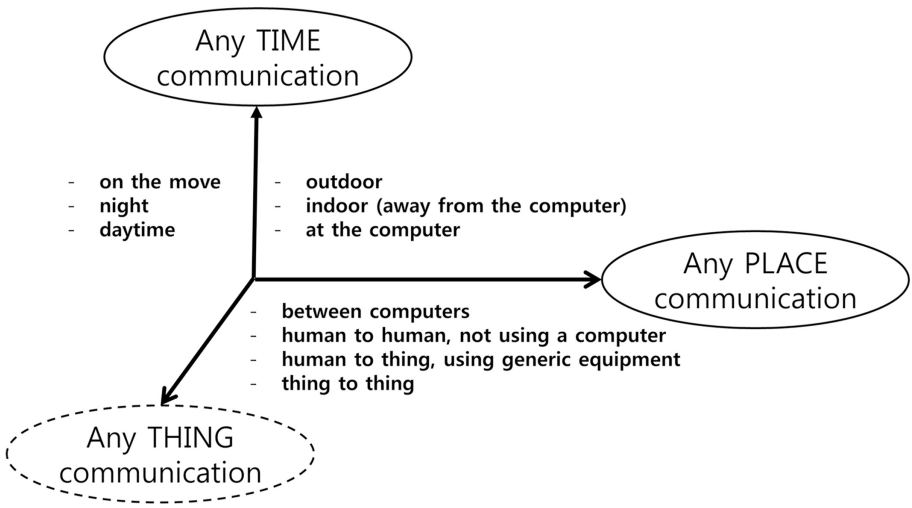


Fig. 1. IoT

real time anywhere, anytime as a concept similar to intelligent machine to machine (M2M) communications [8, 9]. IoT can be defined as ICT platform exchanging all the necessary information between humans and things and between things through the Internet by synthesizing various domestic and international definitions of IoT [10].

3 Proposal

The proposed structure shows from registration procedure to information exchange process through a smart device and CA via each node and gateway in the home IoT environment as shown in Fig. 2. A smart device refers to a smart phone and a tablet PC that can offer information to node through gateway. CA and Gateway exchange information mutually through a safe channel. When registration procedure is finished through mutually safe procedure, limitation was made to the process in which user's request is accepted between the smart device and node through gateway (Table 1).

Table 1. Scheme notation

Notation	Meaning
SN	Node serial number
r_n	Random number
E_{NCA}	Node symmetric key
E_{GCA}	Gate way symmetric key
SK	Session key
TS	Time stamp

3.1 Proposed Protocol

At first, gateway receives node's information, after advertising to node through continuous broadcasting. And then, gateway verifies through decryption, after receiving the encrypted information with node's open key, and sending the information to CA through a safe channel. CA that received information through the safe channel sends the information indicating the node is the legitimate node to gateway. The gateway creates a session key, and uses it for future information exchange. The created session key can prevent re-use attack through continuous renewal. After checking whether the node is legitimate one through CA, the device is registered, after a safe session key is created between the node and smart device via the connection with a smart device. Communications are conducted between the smart device and node, after the node is registered through gateway. The low power process, node, conducts arithmetic operation through gateway, not doing it itself, and thus weight-light and safe communications can be carried out. Also, the concentrated phenomenon of energy consumption of the node that should deliver lots of data can be reduced.

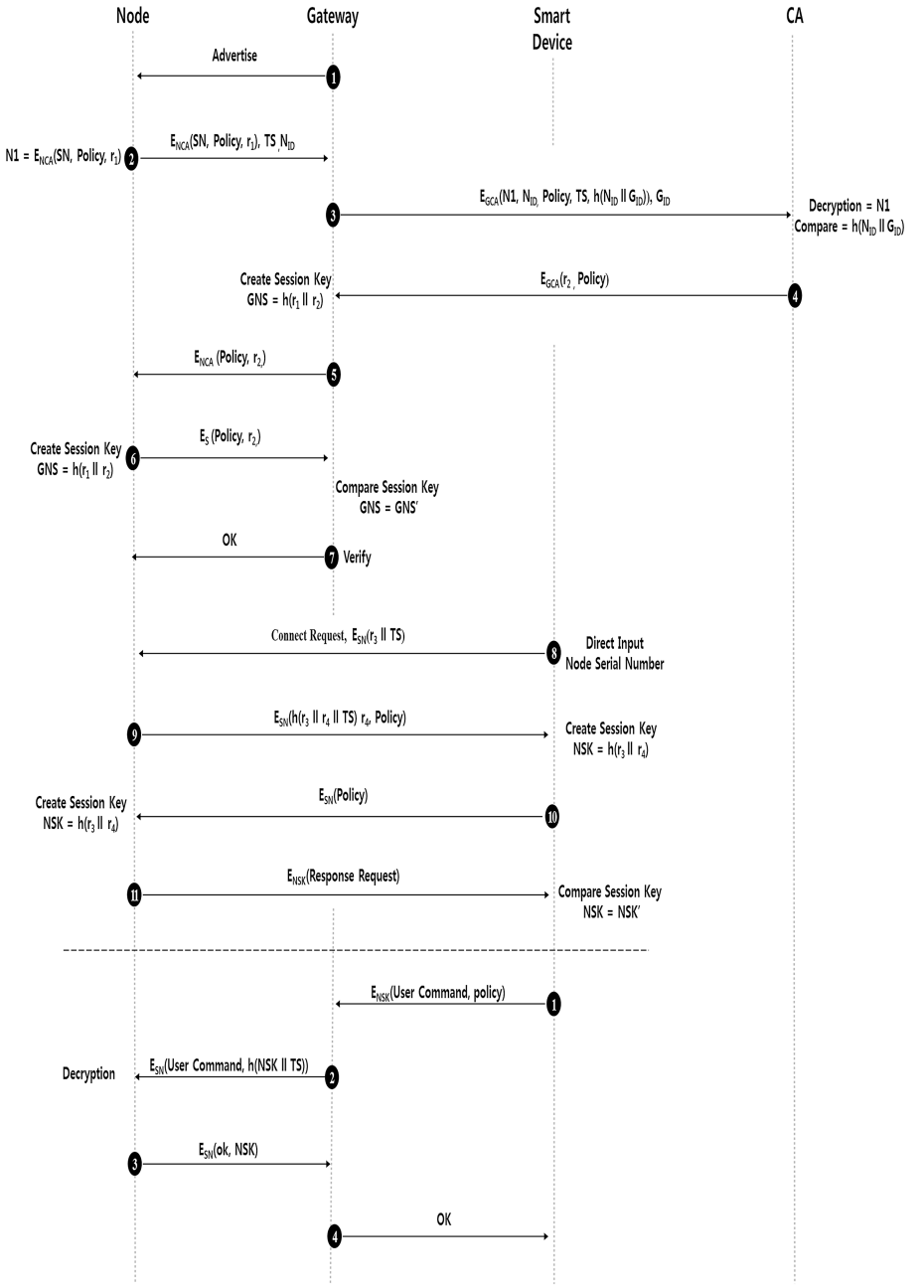


Fig. 2. Authentication protocol

4 Security Evaluation

4.1 Replay Attack

A hacker can try re-transmission attack by acquiring a communications message between a smart device and gateway. However, the communication message between the smart device and gateway includes time stamp value, and therefore, it is safe from the re-transmission of hacker's message.

4.2 Brute Force Cracking

A hacker can try indiscriminate attack to the open key to identify the node SN and Node ID that are encrypted with the open key of the gateway. However, it is safe, because gateway and CA exchange information through the safe channel, and the session key is created, after verification from safe CA.

4.3 Eavesdropping Attacks

A hacker can eavesdrop a communications message between a smart device and gateway, and a communications message between a smart device and node. However, the session key is made by using the information safely received from the random number value R, which changes every time, and time stamp TS value and CA, and thus a hacker cannot identify the node ID or node SN.

5 Conclusion

As the home IoT communications environment is built, the ratio of low specification devices, as well as that of high specification devices, increases. For safe communications in the home IoT environment, the encryption algorithm such as RSA offering message encryption and authorization together is required. However, difficulties are accompanied in using the algorithm to the low specification devices with limited calculation function in the home IoT environment. The security problem of low power devices becomes a security issue, and the damage is conveyed to users. The safe and weight-light user authorization and information exchange process between the node and smart device through gateway offered in this paper is more effective for user damage prevention and the prevention of expansion. The process proposed in this study is indeed expected to contribute to the enhancement of stability and reliability.

References

1. Manyika, J., Chui, M., Bughin, J., Dobbs, R., Bisson, P., Marrs, A.: Disruptive technologies: advances that will transform life, business, and the global economy. McKinsey Global Institute, May 2013
2. Korea Association of Smart Home (2013)

3. Internet of Things & Machine-To-Machine (M2M) Communication Market, Markets and Markets (2012)
4. http://m.it.chosun.com/m/m_article.html?no=2818346
5. ITU Internet Reports 2005: The Internet of Things – Executive Summary, ITU, 2005.11. The internet of things, Commission of The European Communities, September 2008
6. ITU Internet Reports 2005, Internet of Things, ITU, November 2005
7. Internet of Things in 2020, A Roadmap for the Future, EPoSS, September 2008
8. The Internet of Things 2012 New Horizons, IERC (2012). http://www.internet-of-things-research.eu/pdf/IERC_Cluster_Book_2012_WEB.pdf
9. Basic plan for base construction of machine to machine. Korea Communication Commission (2009)
10. Mahoney, J., LeHong, H.: Innovation Insight: The 'Internet of Everything' Innovation Will Transform Business, Gartner, Inc., 3 January 2012

Platform Independent Workflow Mechanism for Bigdata Analytics

Tai-Yeon Ku[1(✉)], Hee-Sun Won[1], and Hoon Choi[2]

[1] Data Management Research Section, Software Research Laboratory, ETRI, Dae-jeon, Korea
{kutai,hswon}@etri.re.kr
[2] Chungnam National University, Dae-jeon, Korea
hc@cnu.ac.kr

Abstract. The present paper allows to platform independent process based Workflow mechanism for big-data analytics. The Bigdata Workflow Tool is to provide a system and a method for data analysis services which is able to select data corresponding to a user requirement and an analysis algorithm which is able to analyze the data and let an analysis service selected by the user among a plurality of analysis services be automatically performed in a big data platform. A system and method for combining a workflow is provided. This paper is directed to a system and method for combining a workflow capable of providing a big data analysis more easily and rapidly and recognizing an analysis flow more easily by combining a model-based workflow.

1 Introduction

Currently, utilization of data which a person has, data which a company has, and data which a public institution has is being limited in the name of systemic and institutional personal information protection. Further, technically, many data analysis algorithms and methods are emerging. However, there is a problem in which a user having data available to utilize these technologies should understand a difficult analysis method in order to actually utilize the technologies. Valuable data is being lost since there is a case in which the user does not know the utilization of data which he/she has. For this, big data platform technology has been developed, and various methods for utilizing the technology are being proposed, but only information technologies (ITs) for advanced users have been offered. In order to solve the problem, a big data ecosystem is being increased based on big data technology while the big data technology is being developed. For this, currently, in order to provide an analysis service using the big data technology, an analysis service fully understanding each of components of an ecosystem and linked with the components in purpose is being developed. Meanwhile, a big data operating analysis environment should provide a service capable of performing a collection of various log data and a service available for an advanced convergence analysis using utilization of an analysis function of various services and an analysis engine framework. Accordingly, since a type and a size of derived log data are exponentially increased as various services are provided, it is necessary to provide an analysis process modeling tool capable of actively coping with an analysis engine

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_38

framework specialized in a service and a change of the service. Recently, due to developments of a big data platform field based on Hadoop, a data market, an analysis algorithm market, etc. are being visualized. Through this, many big data platform technologies are attempting to link data and an analysis algorithm, but there is a problem in that the link thereof is not easy due to absence of a standardized method for this. Accordingly, when providing an analysis service using big data, in a current system, an analysis service fully understanding each of components and linking the components in purpose should be developed. However, in order to understand the components of the ecosystem, a lot of time and effort is required.

2 Overview of Big-Data Workflow Architecture

The present paper is to provide a system of platform independent process based workflow mechanism for big-data analytics. Workflow tool has several function: analysis service creation, play, monitoring. This function performs using web GUI service. Workflow tool defines the selected data and the selected analysis algorithm to a job and include input and output variables and algorithm variables into the defined job to connect the data and the analysis algorithm with the job. The workflow generator may generate analysis information using the received job and transmit the generated analysis information to BDAS platform. The algorithm manager may manage the stored analysis algorithms by being interconnected with a separate database which stores analysis algorithms (Fig. 1).

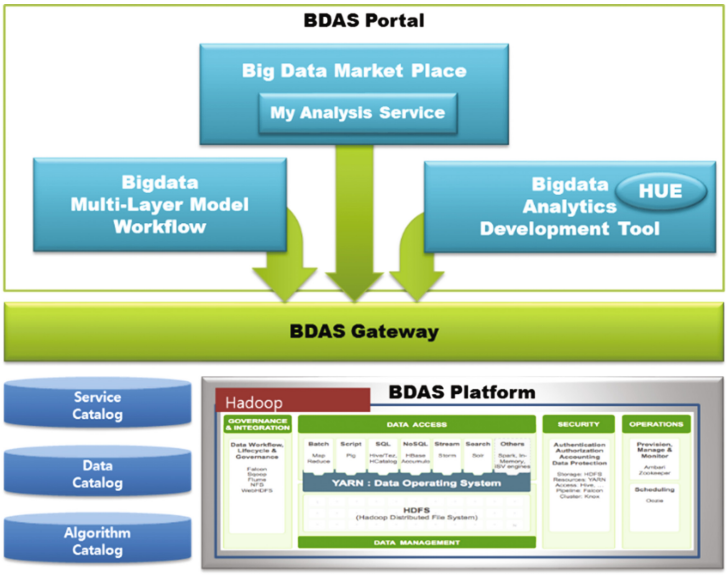


Fig. 1. Platform independent workflow

The analytics workflow may be an analysis processor suite composed of more than one job. For example, more than one job generated through the analysis job generator may be tied to a minimum unit and analysis desired by the user may be then processed. Here, the workflow generator may determine relationship between jobs and manage or modify where current analysis is processing. The workflow scheduler may perform directly to schedule and manage the big data platform and the workflow determined through the workflow generator. The platform interface may be an interface connecting the big data platform and the system for recommending data analysis services. User can set and configure remote data sources through a workflow tool. A workflow tool provides a web-based graphic user interface for developing a service logic and collecting data. User can collect data by developing a workflow diagram in a workflow tool. The present paper is directed to a system and method for combining a workflow capable of providing a big data analysis more easily and rapidly and recognizing an analysis flow more easily by combining a model-based workflow.

According to one aspect of the present paper, there is provided a system for combining a workflow, including: a display module configured to display a workflow model; a memory configured to store a program for combining the workflow model; and a processor configured to execute the program, wherein the processor selects a classification standard for each level of the workflow model in response to a selection input of a user for the classification standard by executing the program, maps one or more workflow models to a node included in the workflow model according to the classification standard for each level, and displays the mapped workflow model, one classification standard is applied to each level, and the workflow model formed as one or more levels is mapped to the node.

3 Platform Independent Workflow Mechanism

There are various technologies and methods in a method for analyzing big data. Among the various methods and technologies, in order to derive various analysis results using a big data ecosystem, technical knowledge and know-how on many ecosystems are required. However, there are temporal and technical limitations in order to understand all of big data analysis technologies, which are rapidly changing, and obtain the best result. For this, this paper may define such ecosystems as a model and provide a reusable structure, and thus a plurality of users may easily and rapidly generate an analysis service through a combination of a puzzle form based on the defined model. Meanwhile, workflow may refer to an illustration of a work processing flow. Further, in the present paper, workflow may refer to a technology of visualizing an analysis process, and illustrating an analysis flow so as to be recognized when developing an analysis service using various eco components of big data. In this case, the present paper may display an analysis flowchart in a graphic user interface (GUI) form using a workflow generation tool. That is, information on a state, etc. of work which is currently being performed, work which is completed, and work which is not processed among work listed on a workflow generation space by performing an analysis may be easily recognized. Further, the analysis result may be displayed in various manners through a visualization model. The workflow model may refer to defining a work flow

which is frequently used in big data analysis as a reusable component. Through a model configured as multi-level, the workflow may be generated in a form such as a puzzle combination suitable for an analysis intention of data (Fig. 2).

Further, through the provided workflow generation tool, a setting error, etc. needed when generating or performing the service of a complex ecosystem may be automatically detected, and a service may be generated using the minimum input needed using a default setting value. Further, even when a function of the ecosystem changes, support for a conventionally generated analysis service may be performed through an update on a related ecosystem model. Accordingly, when generating the service, temporal and economic burden needed for retraining and skill acquisition due to an update of a version of the ecosystem may be minimized. A workflow model may be classified in various forms according to a classification standard. That is, in order to perform an analysis which a user wants, the workflow model may be classified and selected in various forms according to a convenience of the user.

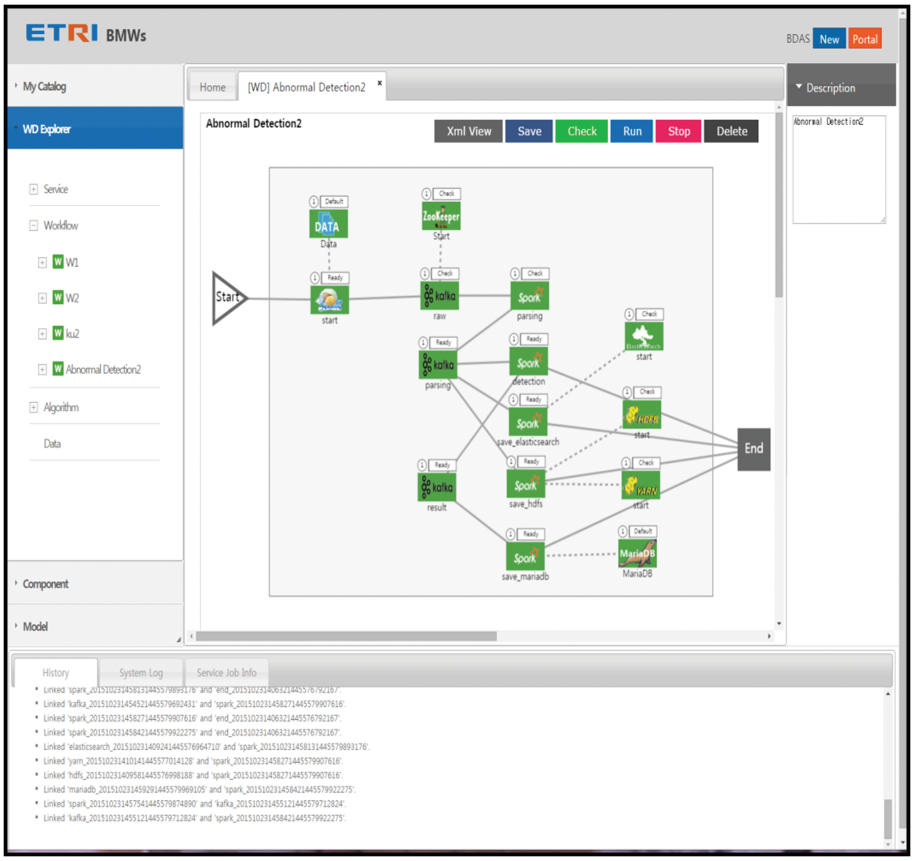


Fig. 2. Workflow authoring tool

4 Use Case

This use-case contains the related data set and analyzing algorithms for spotting and selecting the most probable CCTVs for human screening actions. In developing the processing algorithms and procedures, we needed to investigate available dataset in Korea. A short description about the data set is included in this module package.

This module is the Data Analytics part in the following picture, that shows how the analytics relates with the actual CCTV systems. CCTV Systems are installed in a government regulated authority and they are controlled and monitored in a VMS center. Video Analytics system generates meta data per each event in real time, and the movement meta data is transferred to CAP platform, which becomes one import source of the data analytics. On the other end of the CAP platform, we have external systems, that provides data for accidents, crimes, and weather data (Fig. 3).

Fig. 3. CCTV use case

5 Conclusion

The proposed workflow mechanism for combining the workflow according to this paper may be implemented in a form of a recording medium including a computer program stored in a computer executable medium or a computer executable command. The computer readable medium may be an arbitrary available medium which is able to

be accessed by a computer, and may include a volatile or non-volatile medium, and a separable or non-separable medium. Further, the computer readable medium may include a computer storage medium and a communication medium. The computer storage medium may include the volatile or non-volatile medium, and the separable or non-separable medium implemented using an arbitrary method or technology for storing information such as a computer readable command, a data structure, a program module, or other data. The communication medium may generally include a computer readable command, a data structure, a program module, other data of a modulated data signal such as a carrier wave, or another transmission mechanism, and include an arbitrary information transmission medium.

The method and system of the present paper is described with reference to specific ideas, but some or all of the components or the operations may be implemented using a computer system having a general-purpose hardware architecture. According to this paper, a plurality of users may easily and rapidly generate the analysis service through a combination of a puzzle form based on a defined model by defining the ecosystems as a model and providing a reusable structure. Further, information such as a state of the work which is currently being performed, work which is completed, and work which is not processed, etc. among work listed on a workflow may be easily recognized. The Workflow tool is to provide a system and a method for data analysis services which is able to select data corresponding to a user requirement and an analysis algorithm which is able to analyze the data and let an analysis service selected by the user among a plurality of analysis services be automatically performed in a big data platform.

Acknowledgement. This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT) through the International Cooperative R&D program.

References

1. Ku, T., Won, H., Choi, H.: Service recommendation system for big data analysis. In: ICOIN 2016, October 2015
2. Ku, T., Won, H., Choi, H.: Adaptive cache deploying architecture using big data framework for CDN. In: ICTC 2015, October 2015

Water Surface Simulation Based on Perlin Noise and Secondary Distorted Textures

Hua Li[1,2(✉)], Huamin Yang[1,2], Chao Xu[1], and Yuling Cao[1]

[1] School of Computer Science and Technology,
Changchun University of Science and Technology, Changchun 130022, China
custlihua@gmail.com
[2] National Local Combined Engineering Research Center of Special Film
Technology and Equipment, Changchun 130000, China

Abstract. Simulation of water surface is an important topic in computer graphic. In this paper we propose a fast method to simulate the reflection and refraction of water surface in high quality based on Perlin noise. This method generates the first reflect mapping through mirror reflect. Then the Perlin noise is used to distort the first texture map to generate the secondary reflect mapping which is prospectively projected onto the final surface. Experiment results show that our method can generate high quality reflection with fewer artifacts of reflection.

Keywords: Perlin noise · Reflection · Water simulation · Octave

1 Introduction

Water surface visualization is an element that presented in many natural scenes. Watery surface consistent with the level of detail has been applied in real-time applications such as games, which ignores the wave geometry and sophisticated reflections and refractions. Although many successful techniques have been developed for realistic water surface simulation, these techniques are only appropriate for off-line rendering of scene's sequences or concentrates on modeling the water surface as a texture-mapped plane with simple lighting effects. In this paper, we describe a method that makes it possible to render water surface with high quality reflections and refraction with two stages of textures. Our system can be implemented with vertex and pixel shaders. First we compute the height map using octaves of Perlin noise to model the water waves. The height map is enforced to a mirror reflection which is called the first reflection texture. Then distort the textures according to the height map based on Perlin noise. Next calculate the reflection and refraction based on the distorted textures. Finally, a Fresnel bump mapping is enforced to render per-pixel on the water surface.

2 Previous Work

Early graphics work concentrated on modeling the water surface as a texture-mapped plane with simple lighting effects or establishing a parametric function just animate wave transport [1–3]. The behavior of a volume of water can be described by a set of equations

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_39

developed by Navier-Stokes (fluid dynamics) which can calculate surface of dynamic bump mapped and effects of rays underwater. Jerry [4] uses a statistical wave model based on FFT to represent waves and allows external interactions with the water by Navier-Stokes equations. But this still involves a lot of heavy-weight computation. R. Elliot *et al*. [5] proposed a method for large scale water simulation using Chimera grid embedded in which different regions of interest could be decomposed into varying spatial resolutions of the domain. Water reflection can be viewed in [6, 7]. Hinsinger [8] proposed an algorithm to animate the ocean waves in real time without reflection. Premoze and Ashikhmin [9] proposed a method for wave generation and a light transport approach for complex lighting computing. However, it is a time-consuming and impractical work for real-time applications. Chentanez [10] present a hybrid water simulation method that combines grid based and particles based approaches and achieves real-time performance on modern GPUs. Humphrey describes a simulation based on bump mapping, dudv-displacement and Fresnel effects in real-time where the bump map is distorted over time to get the impression of waves [11]. Truelsen [12] describes shallow water simulation in real-time using bump mapping and UV displacement. Perlin noise [13], a representative example of Lattice gradient noises [14], which is introduced by Ken Perlin, generates noise by interpolating or convolving random values and/or gradients defined at the points of the integer lattice. Perlin noise provides a continuous noise which is similar to random noises found in nature. Bridson *et al.* [15] presented a Perlin-like noise function called curl noise which is used for generating time-varying incompressible turbulent velocity fields. Perlin and Neyret [16] presented flow noise in 2001, which generated time-varying flow textures with swirling and advection. Building the height field for waves using Fast Fourier Transformation (FFT) [4] is popular because it is fast property and reasonable effect. 2D or 3D grid is utilized come up with level of detail (LOD) [17] method or grid projection optimization. Lagae [18] describes a survey on procedural modelling for virtual worlds. Jerry Tessendorf [19] introduced Gerstner Wave to get realism water wave which is popular in movie production.

3 Height Field Generation

3.1 Principle of Perlin Noise

Perlin and Hoffert [20] gave the following definition: *noise is an approximation to white noise band-limited to a single octave*. It generates unstructured signals with any combination of frequencies. 2D Perlin noise determines noise at a point in space by computing a pseudo-random gradient at each of the nearest four vertices on the integer lattice and then doing a splined interpolation. The set of gradient consists of the four corners of lattice cell relative to the sample point which is defined by the directions from the center of the square to its edges. The chosen interpolant should ensure a continuous noise derivative.

The basic function of Perlin noise is Eq. (1) or (2). Diagram of these two functions see Fig. 1 where there is a little difference involved. We utilize Eq. (1) generating 2D Perlin noise as shown in Fig. 2 where multiple octaves of Perlin noise are summed up to get the fractal Perlin noise.

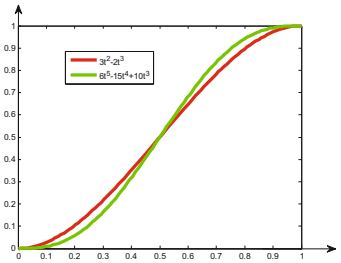


Fig. 1. Perlin noise function.

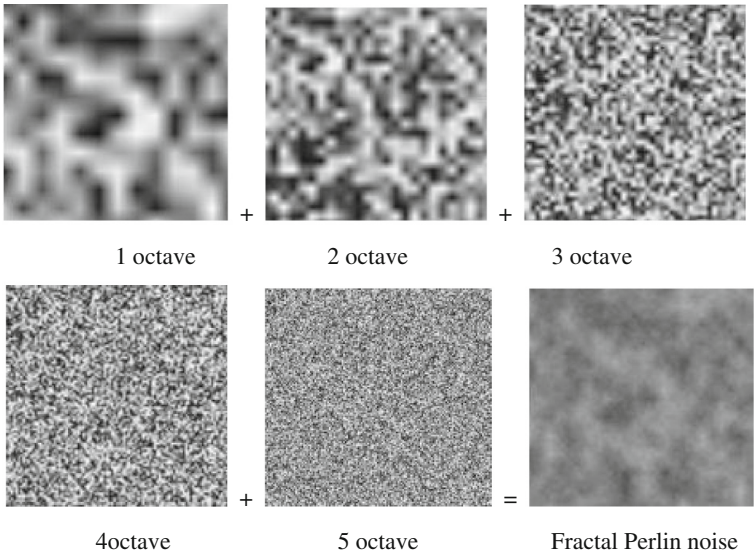


Fig. 2. 2D Fractal Perlin noise are built by summing up multi-octaves.

$$f(t) = 3t^2 - 2t^3 \tag{1}$$

or

$$f(t) = 6t^5 - 15t^4 + 10t^3 \tag{2}$$

In Fig. 2, octaves refer to the frequency of each layer, here the frequency of layer 2 is double of that of layer 1, and layer 3 is double that of layer 2 and so on. This is in fact very close to the random phenomenon in nature and easily the most computationally efficient of the method.

3.2 Generation of Perlin Noise

We compute gradient for an arbitrary sample point by taking dot product of displacement vectors S and their respective gradient vectors G, see Eq. (3). Then we get the gradient of P by interpolate the four gradient values of Q. See Fig. 3.

$$Q = S \cdot G \tag{3}$$

For more intuitive illustrating of the 2D Perlin noise, we generate the random height curve using Eq. (4) and the results in Matlab see Fig. 4, where Eq. (4) is the two dimensions of Eq. (1). We sum up different octaves of Perlin noise to generate a fractal Perlin noise and distort the mirror reflection texture map.

$$\begin{cases} f(x,y) = (3x^2 - 2x^3)(3y^2 - 2y^3) \\ f(x_1, x_2, \ldots x_n) = \prod_{i=1}^{n} (3x_i^2 - 2x_i^3) \end{cases} \tag{4}$$

Perlin noise is added to the mirror texture with gradient and normal of each sample point on the surface.

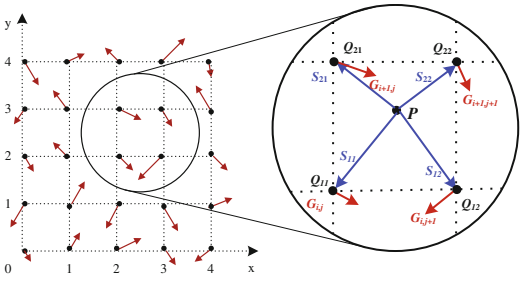


Fig. 3. Calculate gradient for an arbitrary sample *P*.

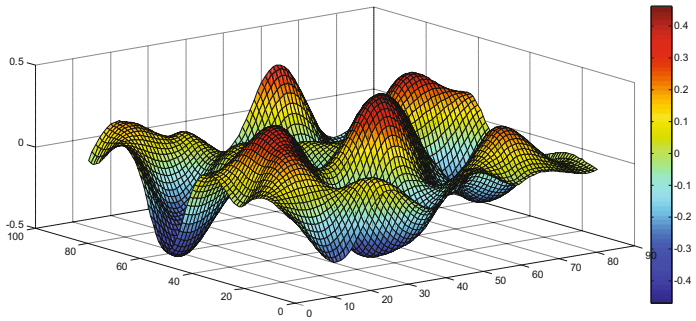


Fig. 4. Radom height simulated by 2D Perlin noise.

4 Generation of Secondary Distorted Texture Mapping

The secondary distorted texture map is obtained by Normal operation. As the gradient of each sample point has been calculated, the normal can derive from the dot product of gradient of x and y direction (see Eq. 5).

$$\vec{n}_P = normalize(\frac{\partial f}{\partial x} \times \frac{\partial f}{\partial y}) \tag{5}$$

where the gradient of x and y are $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ respectively.

4.1 Snell's Law

We calculate reflection and refraction based on Snell's law (Eq. 6). The principle of Snell's law is in Fig. 5, where n_1 and n_2 denote the refraction indices for the two materials respectively: air and water.

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1} = \frac{n_{water}}{n_{air}} \tag{6}$$

In Eq. (6), n_1 and n_2 are the refractive index of the two medias, θ_1 and θ_2 are the included angle between the incident light (and the refraction light) and the interface normal vector. When n_1 is equal to n_2, it is the reflecting situation.

As n_{water} = 1.333 and n_{air} = 1.0003. This gives us Eq. (7):

$$\sin(\theta_{water}) = 0.75\sin(\theta_{air}) \tag{7}$$

Equation (7) is the parameters we used in our method.

From Eq. (6), we have Eq. (8)

$$V_{reflect} = 1 + 2\cos\theta_1 n \tag{8}$$

According to formula (9) and (10):

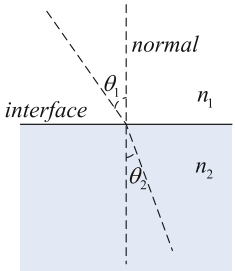


Fig. 5. Principle of surface reflection and refraction.

$$\sin\theta_2 = \left(\frac{n_1}{n_2}\right)\sin\theta_1 = \left(\frac{n_1}{n_2}\right)\sqrt{1-(\cos\theta_1)^2} \tag{9}$$

$$\cos\theta_2 = \sqrt{1-(\sin\theta_2)^2} = \sqrt{1-\left(\frac{n_1}{n_2}\right)(1-(\cos\theta_1)^2)} \tag{10}$$

Refraction calculation is shown as Eq. (11)

$$V_{refract} = \left(\frac{n_1}{n_2}\right)l + \left(\left(\frac{n_1}{n_2}\right)\cos\theta_1 - \cos\theta_2\right)n, \quad n = \{0,1\} \tag{11}$$

4.2 Fresnel Equation

We will represent the reflection and the refraction by textures, and we generate the secondary texture mapping using the color of the water surface which will be the weighted sum between these two textures added with some ambient light as shown in Eq. (12).

$$surface = F * reflection + (1-F) * refraction + light \tag{12}$$

Fresnel term (Fresnel reflectance) is a calculation for determining realistic light reflecting off of a surface. In our method an approximation of the full equation is used as Eq. (13) [11].

$$F = 1 - N \cdot \mathrm{I}_{\mathrm{incident}} \tag{13}$$

We calculate the reflection and refraction using Eqs. (8) and (11) to the first distorted texture map. Finally generate the secondary distorted texture mapping.

5 Implementation

We implemented our method using the GLSL with the environment of Intel(R)Xeon(R) CPU E5620@ 2.40 GHz and NVIDIA GeForce GTX580. We generate water surface with the same octaves of Perlin noise and different resolution of texture mapping, see Fig. 6, where octave is 4 and resolution of secondary texture mapping is at 1024×1024(pixel resolution).

Figure 6(a) shows water surface simulation by our method with 6 octaves of Perlin noise and 512×512 resolution of secondary distorted texture mapping. (b) shows the ocean simulation by Jerry using FFT [4]. Our method focuses on the reflection simulation instead of the water movement, so the surface should be calmer.

As the octaves of Perlin noise determine the water surface simulation, we compare the different octaves generating effects with the same resolutions of texture mapping. See Fig. 7, (a) shows 16 octaves reflection of water surface, (b) and (c) shows 8 octaves

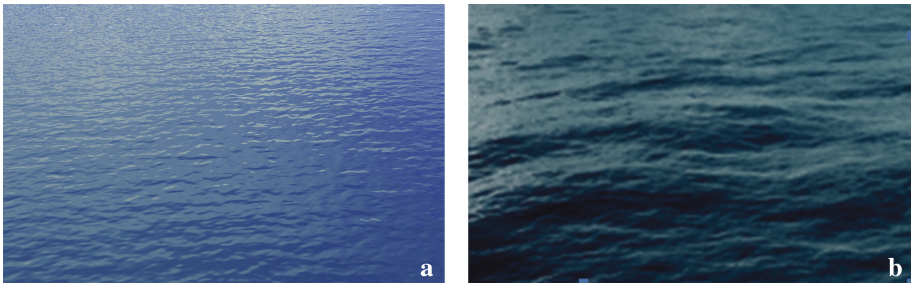


Fig. 6. Water surface simulation. a. our method b. Ocean simulation by Jerry using FFT [4].

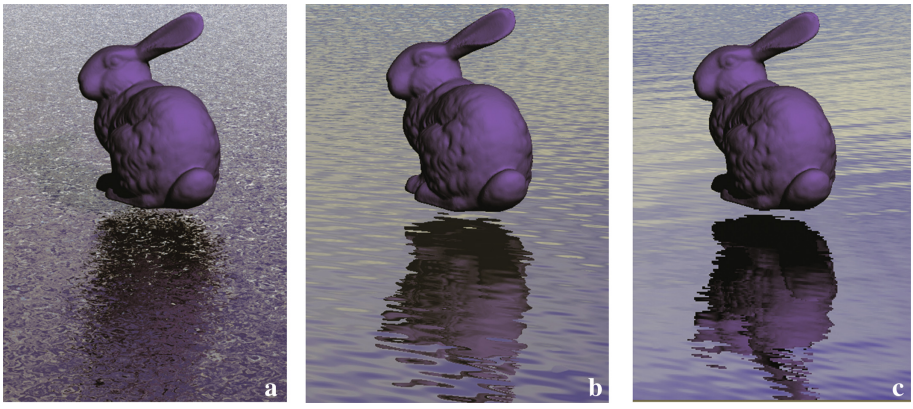


Fig. 7. Water surface simulation with different octaves of Perlin noise and at the same resolution of texture mapping (1024 × 1024). a. 16 octaves, b. 8 octaves, c. 4 octaves.

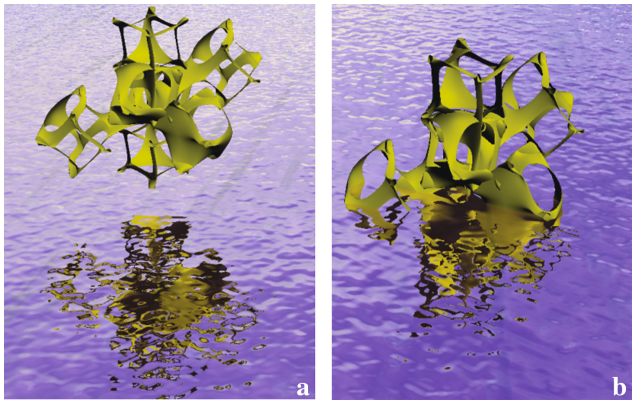


Fig. 8. Reflection and refraction with 4 octaves of Perlin noise at 512 × 512 of texture mapping.

and 4 octaves of water surface simulation respectively. Finally, we generate the water surface with both of reflection and refraction with 4 octaves of Perlin noise and 512 × 512 of texture mapping (see Fig. 8).

We compared our method and Rene's method which is generated based on texture map (see Fig. 9). Rene's method focuses on the shallow water refraction and water surface reflection. But the reflection effects are pretty far from the viewer, such as the shore side which can be used in large landscape simulation. Our method has more detail on the reflection of near objects. Our method has robustness with different resolution of reflection texture map.

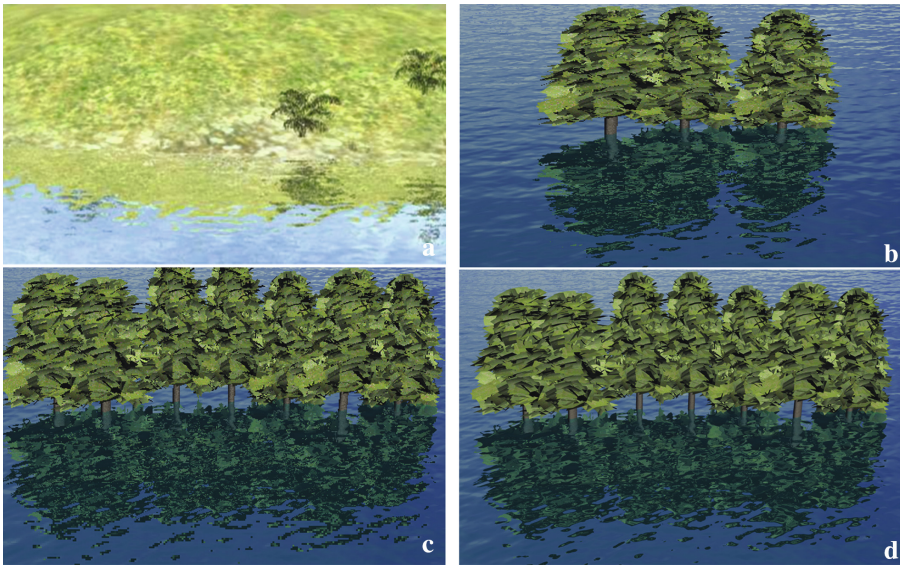


Fig. 9. Comparison of Rene's and Ours. a. Rene's reflection texture at 1024 × 1024 resolution with wave distortion. b. Ours secondary reflection texture at 1024 × 1024 resolution. c. Ours at 256 × 256 resolutions. d. Ours at 1480 × 1480 resolutions.

6 Conclusions

This research aims to simulate water surface with reflection and refraction in real-time. Perlin noise is identified a reasonable procedural noise function to generate random height values. The ability to add complex and intricate detail at low memory and authoring cost is one of the main attractions of Perlin noise. Our method generates good quality effects using the secondary distorted textures with the calculation of reflection and refraction which is used in the application of our movie production projects. Next work we will simulate more visual effect using Perlin noise and develop parallel methods.

Acknowledgements. This work was financially supported by development project of Jilin province science and technology (20140204009GX), Jilin upgrade industrial innovation special fund projects(2016C091) and major scientific and technological plan of Changchun (14KG008, 14KG013).

References

1. Fournier, A., Reeves, W.T.: A simple model of ocean waves. ACM Siggraph Comput. Graph. **20**(4), 75–84 (1986)
2. Peachey, D.R.: Modeling waves and surf. In: ACM Siggraph Computer Graphics, pp. 65–74. ACM, August 1986
3. Schachter, B.: Long crested wave models. Comput. Graph. Image Process. **12**(2), 187–201 (1980)
4. Tessendorf, J.: Simulating ocean water. Simulating nature: realistic and interactive techniques. SIGGRAPH **1**(2), 5 (2001)
5. English, R.E., Qiu, L., Yu, Y., Fedkiw, R.: Chimera grids for water simulation. In: Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 85–94. ACM, July 2013
6. Nishita, T., Nakamae, E.: Method of displaying optical effects within water using accumulation buffer. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 373–379. ACM, July 1994
7. Watt, M.: Light-water interaction using backward beam tracing. ACM SIGGRAPH Comput. Graph. **24**(4), 377–385 (1990)
8. Hinsinger, D., Neyret, F., Cani, M.P.: Interactive animation of ocean waves. In: Proceedings of the 2002 ACM SIGGRAPH, Eurographics Symposium on Computer Animation, pp. 161–166, July 2002
9. Premoze, S., Ashikhmin, M.: Rendering natural waters. Comput. Graph. Forum **20**(4), 189–199 (2001)
10. Chentanez, N., Müller, M.: Real-time simulation of large bodies of water with small scale details. In: Proceedings of the 2010 ACM SIGGRAPH/Eurographics Symposium on Computer Animation, pp. 197–206. Eurographics Association, July 2010
11. Humphrey, B.: Realistic water using bump mapping and refraction, Retrieved June 2006
12. Truelsen, R.: Real-time shallow water simulation and environment mapping and clouds (2007)
13. Perlin, K.: An image synthesizer. ACM Siggraph Comput. Graph. **19**(3), 287–296 (1985)
14. Lagae, A., Lefebvre, S., Cook, R., DeRose, T., Drettakis, G., Ebert, D.S., Zwicker, M.: A survey of procedural noise functions. Comput. Graph. Forum **29**(8), 2579–2600 (2010). Blackwell Publishing Ltd.
15. Bridson, R., Houriham, J., Nordenstam, M.: Curl-noise for procedural fluid flow. ACM Trans. Graph. (TOG) **26**(3), 46 (2007)
16. Bruneton, E., Neyret, F., Holzschuch, N.: Real-time realistic ocean lighting using seamless transitions from geometry to BRDF. Comput. Graph. Forum **29**(2), 487–496 (2010). Blackwell Publishing Ltd.
17. Neider, J., Davis, T., Woo, M.: OpenGL Programming Guide. Addison-Wesley, New York (1993)

18. Smelik, R.M., Tutenel, T., Bidarra, R., Benes, B.: A survey on procedural modelling for virtual worlds. Comput. Graph. Forum **33**(6), 31–50 (2014)
19. Perlin K, Neyret F. Flow noise. In: ACM SIGGRAPH Technical Sketches and Applications, p. 187. (2001)
20. Perlin, K., Hoffert, E.M.: Hypertexture. ACM SIGGRAPH Comput. Graph. **23**(3), 253–262 (1989)

Optimization for Particle Filter-Based Object Tracking in Embedded Systems Using Parallel Programming

Mai Thanh Nhat Truong and Sanghoon Kim(✉)

Department of Electrical, Electronic, and Control Engineering,
Hankyong National University, 327 Jungang-ro,
Anseong-si, Gyeonggi-do, Republic of Korea
kimsh@hknu.ac.kr

Abstract. Object tracking is a common task in computer vision, an essential part of various vision-based applications. After several years of development, object tracking in video is still a challenging problem because of various visual properties of objects and surrounding environment. Particle filter is a well-known technique among common approaches, has been proven its effectiveness in dealing with difficulties in object tracking. In this research, we develop an particle filter-based object tracking method using color distributions as features. Moreover, recently embedded systems have become popular because of the rising demand of portable, low-power devices. Therefore, we also try to deploy the particle filter-based object tracker in an embedded system. Because particle filter is a high-complexity algorithm, we will utilize computing power of embedded systems by implementing a parallel version of the algorithm. The experimental results show that parallelization can increase performance of particle filter when deployed in embedded systems.

Keywords: Object tracking · Particle filter · Parallel computing · Embedded system

1 Introduction

Object tracking has important roles in many vision-based applications, such as industrial inspection, human-computer interfaces, traffic monitoring, surveillance systems. Generally, the goal of object tracking is locating and producing a trajectory record of the objects in image sequences. Among common approaches for object tracking, particle filter is a robust solution. Particle filter has been proven to produce high performance when applied to nonlinear and non-Gaussian estimation problems [1]. This method approximates a posterior probability density of the state, in this case it is the positions of the object in image sequence. The approximation is done by using point mass representations of probability densities, which are called particles. Usually contours, color features, or appearance models are used as particles when applying particle filter to object tracking [1–5]. The color histogram is often used because of its robustness against noise and occlusion, but suffers from illumination changes. This disadvantage can be overcome by using other color spaces which are less sensitive to light conditions.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_40

Although particle filter have been commonly used in recent years, they have considerable disadvantages [6]. One of those is sampling impoverishment. In this phenomenon, high-weight particles have higher chance to be drawn multiple times during resampling, whereas low-weight particles have low chance to be drawn at all. This makes the diversity of the particles decrease after several resampling steps. In the worst case scenario, all particles might be "merged" into a single particle. Sampling impoverishment reduces the performance of tracking process drastically. Another disadvantage of particle filter is computational complexity. The complexity increases as the area of tracked region and the number of particles increase. When the dimensionality of the state space increases, the number of particles required for the sampling increases exponentially.

Embedded systems, however, have limited computing power. This limitation prevents embedded systems from performing complex tasks within reasonable execution time. Fortunately, thanks to the rapid development of semiconductor industry, embedded systems have become more powerful. This research presents a method for utilizing modern hardware architecture of embedded systems in object tracking task. We will use parallel programming to implement particle filter algorithm, for the purpose of increasing the performance of particle filter based object tracking method when deployed in embedded systems.

2 Particle Filter and Color Model

2.1 Particle Filter

The particle filter in this research, which is developed to track movement of objects, is based on Condensation algorithm [1]. Let X_t be the state vector which describes the quantities of a tracked object, and Z_t be the vector which stores all the observations of object movements $\{z_1,\ldots, z_t\}$ up to time t. Usually, the posterior density $p(X_t \mid Z_t)$ and the observation density $p(Z_t \mid X_t)$ are non-Gaussian, which increases the complexity of tracking process.

In particle filter algorithm, the probability distribution, which presents the object movements, is approximated by a weighted particle set $S = \{(s_n, \pi_n) \mid n = 1\ldots N\}$. For each particle, element s represents the hypothetical state, i.e. location, of the tracked object, and a corresponding discrete sampling probability π. The movement of the tracked objects is described by a statistical model. In tracking process, each particle is weighted depending on observations, then N particles are drawn using resampling techniques. The estimation of mean state is calculated at each time step by:

$$E[s] = \sum_{n=1}^{N} \pi_n s_n \tag{1}$$

Because particle filters have ability to model the uncertainty of object movements, it can provide a robust tracking framework. It can consider multiple state hypotheses simultaneously. On the other hand, particle filters are able to produce high accuracy prediction from previous observations, hence it can deal with short-time occlusions and sudden changes in object movements.

2.2 Color Model

In this research, we will combine particle filter and color model by integrating local color distribution into particles. The model is based on [7]. Color distributions are represented by histograms in the RGB color space. Histogram are calculated by $h(x_i)$, that assigns one of the m-bins to a given color at location x_i. Because not all pixels in the window have the same significance to describe the objects, we will apply a weighting scheme to pixels in each window. The farther from the window center a pixel is, the smaller weight it is assigned. The weighting function is as follows:

$$k(r) = \begin{cases} 1 - r^2 & : \quad r < 1 \\ 0 & : \quad otherwise \end{cases} \tag{2}$$

where r is the distance from the window center. Weighting pixels increases the reliability of the color distribution because boundary pixels may belong to the background. It is also possible to use a different weighting function. However, weighting function (2) is used in order to reduce complexity for overall process. The color distribution $p(y) = \{p_u(y)\}_{u=1\ldots m}$ of window at location y is calculated as:

$$P_u(y) = f \sum_{i=1}^{I} k\left(\frac{\|y - x_i\|}{a}\right) \delta[h(x_i) - u] \tag{3}$$

where δ is the Kronecker delta function, a is proportional to the size of the window, I is the number of pixels in windows, and the normalization factor is:

$$f = \frac{1}{\sum_{i=1}^{I} k\left(\frac{\|y - x_i\|}{a}\right)} \tag{4}$$

In object tracking the estimated state is updated at each time interval depending on new observations, hence we need to measure similarity between color distributions. A popular measure is Bhattacharyya coefficient [8]. Considering two densities $p = \{p_u\}_{u=1\ldots m}$ and $q = \{q_u\}_{u=1\ldots m}$, which are calculated color distributions, the coefficient is defined as:

$$\rho[p, q] = \sum_{u=1}^{m} \sqrt{p_u q_u} \tag{5}$$

Larger ρ means the distributions are more similar. If two distributions are identical, we will get ρ = 1. The distance between two distributions are measured by the Bhattacharyya distance:

$$d = \sqrt{1 - \rho[p, q]} \tag{6}$$

The proposed tracker employs the Bhattacharyya distance to update the a priori distribution calculated by the particle filter.

3 Parallel Implementation

Parallelization can utilize the multi-core architecture of modern embedded systems, in which all cores of the processor take part in the calculation process. Parallel implementation can decrease the processing time of these steps, which leads to higher performance due to higher processed frames per second. However, parallelization has its own limitation. Generally in parallel computing, a task is split up into several threads, then the split tasks will be solved separately in each thread. After completing the computation, those threads will communicate with each other to produce final results. In some cases, the time required for communication is higher than the time of solving split tasks. Overall execution time of parallel implementation in these cases may be higher than sequential implementation. This phenomenon is called parallel slowdown. For the purpose of avoiding parallel slowdown, we will use parallel implementation for resampling particles and calculating likelihood between particles and target object. The detailed algorithms for these steps are listed in Table 1. Since there are no data dependencies between particles in these algorithms, parallel threads do not need to communicate with each other in processing. For example, in resampling step, newArr(1), newArr(2), newArr(3), and newArr(4) can be calculated at the same time using four threads and so on, hence reduce the execution time of whole process.

Table 1. Algorithm

Systematic resampling	Likelihood calculation

Create new array newArr for storing output of resampling process
for $i = 1$ to (number of particle)
  $j \leftarrow 0$
  while (probabilities of particle(j)) < (likelihood i with target object)
  $j \leftarrow j + 1$
  endwhile
  newArr(i) ← particle(j)
endfor
	for $i = 1$ to (number_of_particle)
	  Get image region of particle i
	  Calculate likelihood between obtained image region and target object
	endfor

4 Experimental Results

In this section, we present the performance of object tracking method. Both versions of particle filter algorithm, sequential and parallel, are implemented in C++ under Linux operating system. OpenCV library is used for processing video frames. For parallelization, we use OpenMP feature of GNU compiler for the implementation of parallel particle filter. The program is deployed in embedded board Odroid U3, which has a quad-core processor running at 1.7 GHz and 2 GB of RAM. The video used for testing is acquired from [9]. It has 1017 frame, and a resolution of 320 × 240. We will show the efficiency of the particle filter object tracker and the comparison of performance between sequential and parallel implementation.

The tracking process is shown in Fig. 1. It shows the result of the color-based particle filter using 300 particles, in this case the histograms are calculated in the RGB color space using 8 × 8x8 bins. The figure shows that the color-based particle filter yields good performance when applied in object tracking.

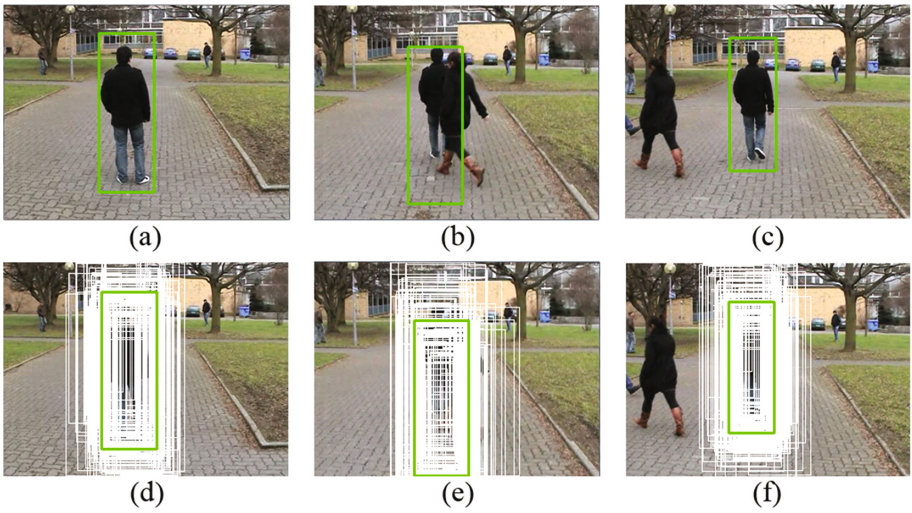


Fig. 1. The object tracker is used to track a walking person in video. Each column shows tracked person (top) and the locations of corresponding particles (bottom).

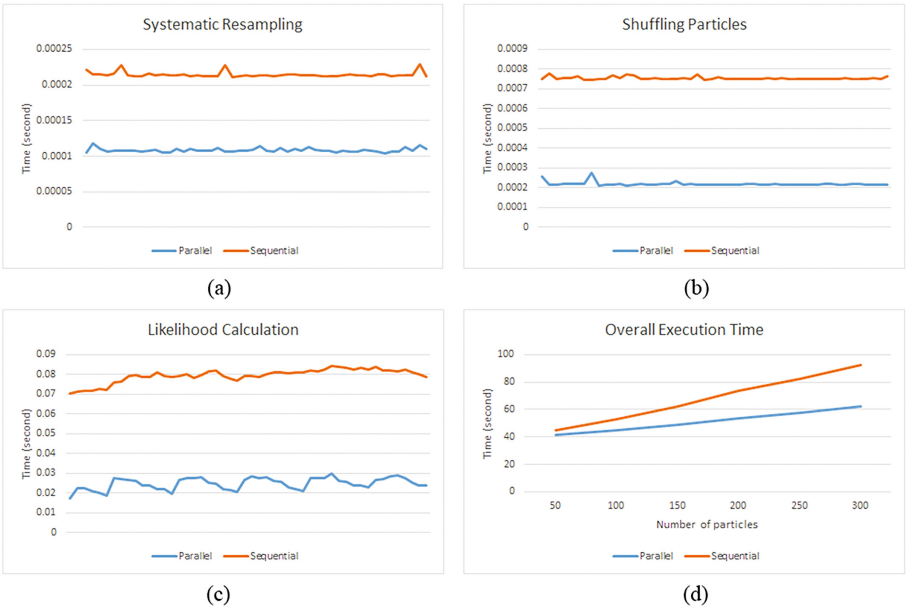


Fig. 2. Comparison of execution time between two implementations.

Figures 2a–c show the comparison in execution time of each step in the first 50 frames between two implementation. The parallel implementation is configured to be executed using four threads. As shown in figures, the parallel implementation is at least two times faster than non-parallel implementation. Figure 2d shows the execution time of both implementations in processing whole video. With 300 particles, parallel implementation took around 62 s to process 1017 frames (around 16 frames per second), while non-parallel implementation took 92 s (11 frames per second). The overall performance is increased by 50 % when applying parallel programming. The results show that, using four threads does not mean the performance will be four times better than using one thread, i.e. sequential programming, because some minor steps in algorithm are processed sequentially. Unfortunately, parallelizing these steps will lead to parallel slowdown.

5 Conclusion

This research aims to increase the performance of particle filter-based object tracking method when the algorithm is deployed in embedded system by using parallel programming. A significant limitation of embedded systems is its computing power, hence it is hard to use embedded systems for high-complexity computation. After several years of development, now embedded systems have become more powerful. At the present modern systems have high memory and multi-core architecture. However, most of the applications for embedded systems do not utilize this advantage. In this research, we have implemented parallel particle filter for object tracking in order to utilize the multi-core architecture. This implementation is deploy in Odroid, an embedded board whose processor has four cores. With 300 particles, the object tracker with sequential implementation can process 11 frames per second, whereas parallel implementation can process 16 frames per second while retaining the tracking accuracy. The experimental results show that multi-core embedded systems can produce higher performance if the hardware is used at its maximum potential.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01057518).

References

1. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. Int. J. Comput. Vis. **29**, 5–28 (1998)
2. Isard, M., Blake, A.: Icondensation: Unifying low-level and high-level tracking in a stochastic framework. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 893–908. Springer, Heidelberg (1998). doi:10.1007/BFb0055711
3. MacCormick, J., Blake, A.: A probabilistic exclusion principle for tracking multiple objects. In: Proceedings of the International Conference on Computer Vision, pp. 572–578 (1999)

4. Ying, W.: Robust visual tracking by integrating multiple cues based on co-inference learning. Int. J. Comput. Vis. **58**, 55–71 (2004)
5. Ko, S.-S., Liu, C.-S., Lin, Y.-C., Khan, O.Z., Balch, T., Dellaert, F.: A rao-blackwellized particle filter for eigentracking. In: Proceedings of the IEEE Conference on Computer Vision Pattern Recognition, vol. 2, pp. 980–986 (2004)
6. King, O., Forsyth, D.,A.: How does CONDENSATION behave with a finite number of samples? In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1842, pp. 695–709. Springer, Heidelberg (2000). doi:10.1007/3-540-45054-8_45
7. Nummiaro, K., Koller-Meier, E., Gool, L.: Object tracking with an adaptive color-based particle filter. In: Gool, L. (ed.) DAGM 2002. LNCS, vol. 2449, pp. 353–360. Springer, Heidelberg (2002). doi:10.1007/3-540-45783-6_43
8. Aherne, F.J., Thacker, N.A., Rockett, P.I.: The Bhattacharyya metric as an absolute similarity measure for frequency coded data. Kybernetika **32**, 1–7 (1997)
9. Tracking Dataset. http://cmp.felk.cvut.cz/∼vojirtom/dataset/. (Accessed 01 May 2016)

Generalized Multi-linear Mixed Effects Model

Chao Li(✉), Lili Guo, Zheng Dou, Guangzhen Si, and Chunmei Li

Group for Future Technology, Department of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China
{guolili, douzheng, siguangzhen, lichunmei}@hrbeu.edu.cn

Abstract. Recently, many applications tend to find common and distinctive features from a group of datasets, of which distributions and structures are generally various. However, most existing methods can just cope with specific problems with fixed distributions and structures. In this paper, a more flexible framework for multi-block data learning is proposed. There are mainly two advantages compared with previous methods: (a) the proposed method can extract global common, local common, and distinctive features automatically; (b) various distributed datasets can be processed simultaneously as long as distributions are in exponential family. The results of numerical experiments demonstrate that the proposed method outperforms conventional methods for recommendation system problems.

Keywords: Tensor decomposition · Multi-block learning · Recommendation system

1 Introduction

Over the years, many methods were developed to extract common and distinctive features from a group of datasets, which are assumed to share common features among datasets, such as fMRI data and its corresponding stimulus in neuroscience [1], multi-relational learning and recommendation system in social data analysis [2], and multi-camera image processing [3]. Compared to conventional "single dataset" problems, there are mainly two challenges for multi-block data learning.

1. Distributions among datasets can be different. For example, In MovieLens, one part of data follows Bernoulli distribution, while another part follows Binomial distribution.
2. Many methods subjectively classify latent features into two subsets, i.e. common features and distinctive features [4]. However, it is obvious that some latent features only owned by part of datasets. Thus how to efficiently recognize features as global common, local common, and distinctive ones makes another challenge.

To overcome these challenges, we proposed a novel method to model multi-block data with more flexible structures. Specifically, since distributions of datasets are usually contained in exponential family, we extended Generalized Linear Model (GLM) [5] into multi-block data learning. Moreover, we developed a novel method to recognize latent features as correct groups by introducing auxiliary modes.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_41

2 Framework

In the model, we use a generalized CANDECOMP/PARAFAC (CP) decomposition to fit each individual dataset. For each dataset $\mathbf{X}_k$, $k = 1, \ldots, K$, we have

$$\mathbf{X}_k \approx f_k\left([[\mathbf{U}_{\{i\in\mathbb{I}_k\}}]]\right), \tag{1}$$

where $\mathbb{I}_k$ denotes index set corresponding to $\mathbf{X}_k$, $\mathbf{U}_{\{i\in\mathbb{I}_k\}}$ denotes a set of latent matrices $\{\mathbf{U}_i | i \in \mathbb{I}_k\}$, $f_k(\cdot)$ represents an element-wise non-linear mapping, and $[[\cdot]]$ denotes kruskal product operator. It can be found from Eq. (1) that each dataset $\mathbf{X}_k$ is fitted by a set of latent matrices $\mathbf{U}_{\{i\in\mathbb{I}_k\}}$, and they own the same features (columns of latent matrices) among each other to share the common information.

2.1 Generalized Multi-linear Model

As mentioned above, the first challenge of multi-block data learning is that different datasets may follow different distributions. Thus, conventional matrix/tensor decomposition cannot simultaneously fit them well. In our model, we extend the idea of Generalized Linear Model (GLM) to fit heterogeneous multi-datasets. In the scenario of matrix, Principal Component Analysis for the Exponential Family (PCA-EF) [6] and Collective Matrix Factorization [7] have proved the efficiency of the idea by lots of applications, and such an idea was also developed in single-tensor factorization. Hence, in this paper we extend this idea to multi-tensor case. By convenience, we assume an individual 3rd-order tensor $\mathbf{X} \in \mathcal{R}^{I\times J\times K}$ and its estimation $\mathbf{Z} \in \mathcal{R}^{I\times J\times K}$, we consider the Bregman divergence taking the form

$$B_F(\mathbf{Z}\mathbf{X}) = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} F_{ijk}\left(z_{ijk}\right) + F^*_{ijk}\left(x_{ijk}\right) - z_{ijk}x_{ijk}, \tag{2}$$

where $F : \mathcal{R}^{I\times J\times K} \to \mathcal{R}^{I\times J\times K}$ denotes an element-wise non-linear mapping in which $F_{ijk} : \mathcal{R} \to \mathcal{R}$ is the mapping corresponding to the (i, j, k) th element, and $F^*_{ijk} : \mathcal{R} \to \mathcal{R}$ denotes the convex conjugate of $F_{ijk}(\cdot)$ which is given by $F^*_{ijk}(z) = \sup_{\alpha}\left[\alpha z - F_{ijk}(\alpha)\right]$. According to [8], the derivative of $F(\cdot)$ can be considered as "link function" which represents a mapping from multi-linear to a more complex structure. Table 1 provides specific forms of $F_{ijk}(\cdot)$ associated to four common distributions where $\dot{F}_{ijk}(\cdot)$ denotes the derivative of $F_{ijk}(\cdot)$.

It is worth noting that different distributions of datasets do not influence the prior of common features in the model. The only difference is the choice of link function. As long as distributions belonging to exponential family, we can use Eq. (2) to fit heterogeneous datasets.

Table 1. Various functions $F_{ijk}(\cdot)$ for four members of the exponential family

DIS	$F_{ijk}(x)$	$\dot{F}_{ijk}(x)$	$F^*_{ijk}(x)$
Normal	$x^2/2$	x	$x^2/2$
Bernoulli	$\log(1+e^x)$	$1/(1+e^{-x})$	0 if x = 0, 1
Binomial	$L \cdot \log(1+e^x)$	$L/(1+e^{-x})$	$x\log\left(\frac{x}{L}\right) + (L-x)\log\left(1-\frac{x}{L}\right)$
Poisson	e^x	e^x	$x\log(x) - x$ if x > 0; 0 if x = 0

2.2 Auxiliary Mode and Feature Selection

The second challenge arises since the most existing methods believe that common features are owned by all the datasets, and remaining features only owned by individual datasets, respectively. However, the fact is usually not like this. Thus modelling features not only as common and distinctive features, but also "local common" features are more suitable for real applications. To our best knowledge, the conception of "local common" was firstly proposed in [9], and they use hierarchical extraction to estimate local common features. But the complexity exponentially increased for large group of datasets. In this paper, we impose auxiliary modes in the model. By using auxiliary modes, we can easily classify features into global common, local common, and distinctive features.

Generally, it is straightforward to model a matrix as 2nd-order tensor. In this case, CP decomposition is equivalent to ordinary matrix low-rank factorization. However, if we model the matrix as a 3rd-order tensor, the CP decomposition becomes matrix tri factorization, and such operation can be extended to higher-order situations. Specifically, assume a 3rd-order tensor $\mathbf{X} \in \mathcal{R}^{I\times J\times K}$, but we use a 4th-order CP model to fit it, then we have

$$\mathbf{X} \approx [[\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4]] = \sum_i u_{4,i}[[\mathbf{U}_{1,i}, \mathbf{U}_{2,i}, \mathbf{U}_{3,i}]], \tag{3}$$

where $u_{4,i}$ denotes the i th element of vector $\mathbf{U}_4$, and $\mathbf{U}_{k,i}$, $k = 1, \ldots, 3$ denotes the i th column of the k th matrices. As auxiliary mode, $\mathbf{U}_4$ can be considered as weight acting on other features. Thus, the value of $\mathbf{U}_4$ controls whether the data own features or not. If there are more than one dataset, the absolute value of auxiliary mode decides the dataset owning the feature or not. If one feature is owned by all datasets, then we call it as global common feature, while it's owned by parts of datasets, then we call it as a local common or distinctive feature.

In order to obtain more structural features, it is straightforward to impose sparsity on auxiliary modes. It is known that L_1 norm is not a bad choice for modelling. But L_1 norm is hard to use in algorithms since its non-smoothness. To achieve a smooth regularization of sparsity, we further add one auxiliary mode, and impose L_2 norm regularization on both of them. In this case, we have

$$\mathbf{X} \approx [[\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_4, \mathbf{U}_5]] = \sum_i u_{4,i} u_{5,i} [[\mathbf{U}_{1,i}, \mathbf{U}_{2,i}, \mathbf{U}_{3,i}]]. \tag{4}$$

Furthermore, it is known that L_2 regularization is equivalent to imposing Gaussian prior on latent matrices. If we suppose $u_{4,i}$ and $u_{5,i}$, $\forall i$ are independent normally distributed variables with variances σ_1^2 and σ_2^2, the distribution of their product $w_i = u_{4,i} * u_{5,i}$, $\forall i$ is given by [10]

$$p(w_i) = \frac{K_0\left(\frac{|w_i|}{\sigma_1 \sigma_2}\right)}{\pi \sigma_1 \sigma_2}, \tag{5}$$

where $K_0(\cdot)$ denotes a modified Bessel function of the second kind, and $\mathbf{w} = [w_1, \ldots, w_R]$ are mutually independent. It can be found from Eq. (5) that the regularization on $\mathbf{w}$ tends to the sparsity structure like Bridge regulation. However, unfortunately, this regularization is also non-convex like Bridge regulation.

Indeed, we can impose any numbers of auxiliary models into our model, but it shown from simulations that much more auxiliary modes can lead to slow convergence rate. Thus one or two auxiliary modes are recommended to get the sparse structure for global, local, and distinctive feature extraction.

3 Objective and Algorithm

According to the framework introduced in Sect. 2, we can achieve an objective function for our model. Assume that there are K datasets and I latent matrices, then the objective function is given by

$$L(\mathbf{U}_1, \ldots, \mathbf{U}_I) = \sum_{k=1}^{K} B_{F_k}\left([[\mathbf{U}_{\{i \in \mathbb{I}_k\}}]]\mathbf{X}_k\right) + \sum_{i=1}^{I} \lambda_i P_i(\mathbf{U}_i) \tag{6}$$

where $B_{F_k}(\cdot \parallel \cdot)$, $k = 1, \ldots, K$ denotes Bregman divergence as Eq. (2), $\mathbb{I}_k$ denotes index set that is used to represent which latent matrices $\mathbf{U}_i$ belong to $\mathbf{X}_k$, and $\mathbb{I}_k$, $k = 1, \ldots, K$ can be obtained by Matlab operator, namely $\mathbb{I}_k = \text{find}(\mathbf{A}(k, :) == 1)$. $P_i(\cdot)$ and λ_i, $i = 1, \ldots, I$ denotes L1/L2 regularization items and the corresponding tuning parameters, respectively. It should be noted that distributions of datasets are uniquely determined by non-linear mapping $F_k(\cdot)$, $k = 1, \ldots, K$, and we usually choose one value of λ for all tuning parameters in practice. To search the optimal point of Eq. (6), gradients for latent matrices are given by

$$\nabla L_{\mathbf{U}_i} = \sum\nolimits_{k,\, s.t. a_{ki}=1} \left[\left(\dot{F}_k\left([[\mathbf{U}_{\{k \in \mathbb{I}_k\}}]]\right) - \mathbf{X}_k\right) * \mathbf{W}_k\right] \left(\sum_{j=1}^{i} a_{kj}\right) \left[\bigodot_{l \in \mathbb{I}_k, l \neq i} \mathbf{U}_l\right] + \lambda_i \dot{P}_i(\mathbf{U}_i),\ \forall i, \tag{7}$$

where $[\cdot]_{(i)}, \forall i$ means unfolding of a tensor along the i th mode, $\dot{F}_k(\cdot)$ and $\dot{P}_i(\cdot)$, $\forall i, k$ denotes element-based derivative of $F_k(\cdot)$ and $P_i(\cdot)$ respectively, a_{ki} denotes the (k, i) th element of $\mathbf{A}$, $\mathbf{W}_k$, $\forall k$ denotes a binary tensor in which '1' represents the corresponding element is observed and '0' means not, $*$ denotes Hadamard product, and the definition of operator $\odot$ follows the literature [11].

After achieving gradients, any gradient-based method can be implemented to search the solution. According to [12], Conjugated Gradient (CG) and Newton-like methods such as L-BFGS are recommended to obtain a higher convergence rate. We name this proposed method in this paper as Generalized Multi-linear Mixed Effects Model (GMMM).

4 Numerical Experiment

In this part, we use GMMM to predict users' preference from datasets. Such a problem is usually found in recommendation systems. From the view of machine learning, this problem is actually equivalent to data completion. In this experiment, we use GMMM to predict rating information from MovieLens-100 k. This data consists of 100000 ratings (1–5) from 943 users on 1682 movies, simple information for the users (age, gender, occupation, etc.) and movies (title, date, genre, etc.). We use these information to generate three matrices (User-Movie(UM), Movie-Genre(MG), User-Profile(UP)), assume that elements in UM follow Binomial distribution, and another two datasets follow Bernoulli distribution. Further, we randomly remove part of ratings according to specific percentage, and use observations to predict missing ratings. Additionally, Fast MMMF [13] and CMF [7] are chosen for comparison.

We choose CG as optimization method, initial rank $R = 30$, and $\lambda = [28, 28, 28, 2, 2, 2, 2]$ as tuning parameters for each latent matrices. In Fast MMMF and CMF, we choose the optimal parameters by 10-unfold cross validation. Mean Absolute Error (MAE) for each method under 10 times run is shown in Table 2.

Table 2. MAE of missing rating prediction with different missing percentage

Missing percentage (%)	10	20	50
Fast MMMF	0.81 ± 0.01	0.79 ± 0.01	0.78 ± 0.00
CMF	0.69 ± 0.01	0.69 ± 0.01	0.74 ± 0.01
GMMM	$\mathbf{0.67 \pm 0.01}$	$\mathbf{0.68 \pm 0.00}$	$\mathbf{0.70 \pm 0.00}$

It is shown from Table 2 that GMMM obtains the best performance compared with other two methods. The reason CMF and GMMM outperform Fast MMMF is that CMF and GMMM can exploit the information from all three datasets, while Fast MMMF as a classical collaborative filtering method just can only use User-Movie matrix. Meanwhile the reason GMMM outperforms CMF is that GMMM exploits inherent structure of belongs of features in datasets.

5 Conclusion

In this paper, we introduced a novel model to extract global common, local common, and distinctive features from a group of datasets. The proposed method can handle multi datasets of which not only distributions but also structures of dependence are various. Specifically, to solve the problem of heterogeneous structure, the idea of GLM was extended into multi-block data learning. Furthermore, we developed a conception called auxiliary mode. By imposing auxiliary modes, we can automatically recognize latent features into global common, local common, and distinctive ones. Finally, numerical results showed the proposed method outperforms conventional methods in recommendation system.

References

1. Mitchell, T.M., et al.: Predicting human brain activity associated with the meanings of nouns. Science **320**(5880), 1191–1195 (2008)
2. Nickel, M.: Tensor factorization for relational learning. Diss. lmu (2013)
3. Chen, C.-Y., Grauman, K.: Inferring unseen views of people (2014)
4. Zhou, G., Cichocki, A., Xie, S.: Common and individual features analysis: beyond canonical correlation analysis, Arxiv preprint (2012)
5. Nelder, J.A., Baker, R.J.: Generalized linear models. Encyclopedia of Statistical Sciences (1972)
6. Collins, M., Dasgupta, S., Schapire, R.E.: A generalization of principal components analysis to the exponential family. In: Advances in Neural Information Processing Systems (2001)
7. Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (2008)
8. Bouchard, G., Yin, D., Guo, S.: Convex collective matrix factorization. In: Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (2013)
9. Lfstedt, T., Hoffman, D., Trygg, J.: Global, local and unique decompositions in OnPLS for multiblock data analysis. Anal. Chim. Acta **791**, 13–24 (2013)
10. Aroian, L.A.: The probability function of the product of two normally distributed variables. Ann. Math. Stat. **18**, 265–271 (1947)
11. Zhao, Q., Zhang, L., Cichocki, A.: Bayesian CP factorization of incomplete tensors with automatic rank determination. IEEE Trans. Pattern Anal. Mach. Intell. **37**(9), 1751–1763 (2015)
12. Acar, E., Dunlavy, D.M., Kolda, T.G.: A scalable optimization approach for fitting canonical tensor decompositions. J. Chemometr. **25**(2), 67–86 (2011)
13. Rennie, J.D.M., Srebro, N.: Fast maximum margin matrix factorization for collaborative prediction. In: Proceedings of the 22nd International Conference on Machine Learning. ACM (2005)

Anomaly Detection of Spectrum in Wireless Communication via Deep Autoencoder

Qingsong Feng, Zheng Dou(✉), Chunmei Li, and Guangzhen Si

Group for Future Technology, Department of Information and Communication Engineering, Harbin Engineering University, Harbin, Heilongjiang, China
{fengqingsong, douzheng, lichunmei, siguangzhen}@hrbeu.edu.cn

Abstract. Anomaly detection has been a typical task in many fields, as well as spectrum monitoring in wireless communication. In this paper, we apply a deep-structure autoencoder neural network to spectrum anomaly detection, and the time-frequency diagram is used as the feature of the learning model. In order to evaluate the performance of the model, the accuracy of the output is considered. We compare the performance of both our proposed model and conventional one-layer autoencoder. The results of numerical experiments illustrate that our model outperforms the one-layer autoencoder based method.

Keywords: Anomaly detection · Spectrum monitoring · Wireless communication · Autoencoder

1 Introduction

Anomaly detection aims at recognizing behaviors that do not confirm to the expect behavior, and such behaviors are defined as anomalies [1]. It has been a significant problem in many kinds of fields and applications such as fraud detection of credit card [2], intrusion detection of network [3] and many other domains.

People are in demand for information with the development of information technology. The occurrence of abnormal events is inevitable along with communication and transmission of information in a wireless communication network. Accordingly, the problem how to detect anomalies in a wireless communication network should be addressed, which is of great significance to improve the quality of communication.

Researches on anomaly detection have made a great progress in image processing and speech recognition. For instance, a multi-scale non-parametric approach was proposed to detect and localize anomalies in video surveillance [4]. In order to solve the problem of sub-pixel anomaly detection in hyper-spectral images, a detection method based on maximum anomaly value was proposed [5]. In [6], Least-squares probability classifier was applied to non-parametric based model, to solve the anomaly detection in static and sequential data. A unsupervised approach based on a denoising autoencoder with bidirectional LSTM recurrent neural networks for automatic acoustic novelty detection was presented [7]. Later, this model was improved by non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection and approached a significant performance [8].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_42

This paper puts forward a novel purely unsupervised approach for signal anomaly detection of spectrum in wireless communication based on deep autoencoder. The signal spectral features are processed by the autoencoder, which acts as a one-class classifier. Our approach relies on the reconstruction error to judge whether the signal is anomaly or not. We compare the results of conventional one-layer network with our proposed two-layers network, and conclude that the performance of proposed two-layers network is superior than conventional one-layer network.

2 Model of Autoencoder

An autoencoder neural network is a kind of unsupervised learning method that applies backpropagation algorithm, setting the target outputs to be equal to the inputs. A basic autoencoder consists of three layers, including input layer, hidden layer and output layer. It is often used in deep neural networks to find the common representation of the inputs. In response to an input $x \in R^n$, the hidden layer representation $h(x)$ is

$$h(x) = f\left(W_1^{m\times n}x + b_1\right),$$

where $f(\cdot)$ is a non-linear activation function. Generally, we choose a logistic sigmoid function $f(x) = 1/(1+\exp(-x))$ as the activation function. $W_1^{m\times n}$ represents the weight matrix of the hidden layer, and $b_1 \in R^m$ denotes the bias vector of the hidden layer.

The output layer representation $h(x)$, which equals the reconstruction value $\tilde{x}$ is

$$\tilde{x} = f(W_2^{n\times m}h(x) + b_2),$$

where $W_2^{n\times m}$ represents the weight matrix of output layer, $b_2 \in R^n$ denotes the bias vector of output layer.

Thus given the input set $\boldsymbol{x}$, we have the reconstruction error as

$$\varepsilon = \sum_{x\in\boldsymbol{x}} \|x - \tilde{x}\|^2$$

In order to get the minimum reconstruction error, stochastic gradient descent algorithm is used to train the neural network. In this paper, we conduct our experiment on two kinds of neural network structures, conventional one-layer autoencoder network and proposed two-layers autoencoder network. The neural networks are developed and trained through Neural Network Toolbox in MATLAB.

3 Experiment and Results

Our experiment is conducted based on an RTL-SDR device. RTL-SDR is a kind of DVB-T dongle that can provide a critical component of a cheap SDR (Software Defined Radio), since the chip allows transferring the raw I/Q samples to the host.

In our experiment, RTL-SDR device is used to collect the real-time data from electromagnetic environment. The data will be first pre-processed by our system. After that, the pre-processed data will be used to train and test our model. The process diagram of our model is shown in Fig. 1, and the specific realization will be illustrate after that.

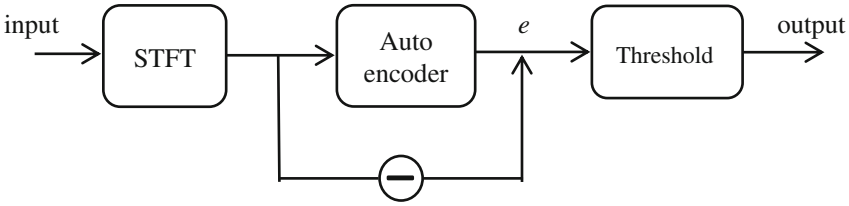


Fig. 1. Procedure of anomaly detection using deep autoencoder network

3.1 Threshold Determination

In order to distinguish anomalies from normal, a threshold is often applied to transform the reconstruction errors into a binary signal. The selection of the value of threshold is a trade-off between wrong alarm probability and missing alarm probability. Typically, the threshold is defined as the median of a sequence of reconstruction errors. With a coefficient α, the threshold can be written as

$$\beta = \alpha * median(e_1, e_2, \ldots, e_N),$$

where $\{e_1, e_2, \ldots, e_N\}$ represents the reconstruction error of different data samples. As to find a threshold that makes our model have a better performance, an enumeration method is used in this paper instead of the conventional median approach. The range of our enumeration is from the maximum reconstruction error of the samples without anomalies to the minimum reconstruction error of the samples with anomalies.

3.2 Data Pre-process

In order to investigate the frequency of signal at each time point, we will take a Short-Time Fourier Transform (STFT) method to pre-process the input data samples at first. STFT is defined as

$$X_n\left(e^{j\omega}\right) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)e^{-j\omega m}.$$

In our experiment, the collected data is pre-processed with 512 points STFTs. The sampling rate equals 3 M symbols per second, thus, we can get the frequency resolution is about 6 kHz. Futhermore, a Hanning window function is selected as the window function in our data pre-processing procedure.

3.3 Network Structure

The experiment is conducted in two different structures of neural network, one of which is conventional one-layer autoencoder network and the other is our proposed two-layers autoencoder network. The parameters of the network are shown in Tables 1 and 2.

Table 1. Parameters of conventional one-layer network

Parameter	Value
Hidden size	20
Max epochs	200
L2Weight regularization	0.004
Sparsity regularization	4
Sparsity proportion	0.15

3.4 Model Selection

The selection of hidden size is conducted by cross-validation method. First, we upset the train data samples and divide them into five segments of equal length. Then we choose one segment randomly as test data samples and the others as training data samples to train and test the neural network. The reconstruction error is recorded as assessment index to choose a better hidden size parameter. In our experiment, the range of hidden size in first layer is from 5 to 120 while that in second layer is from 5 to 100, and the step size of hidden size is 10. The figure of contour line is shown in Fig. 2.

Table 2. Parameters of proposed two-layers network

Parameter	Value of 1st layer	Value of 2nd layer
Hidden size	60	60
Max epochs	200	200
L2Weight regularization	0.004	0.0004
Sparsity regularization	4	0.0004
Sparsity proportion	0.15	0.1

3.5 Experimental Procedure

In this paper, Additive White Gaussian Noise (AWGN) is selected as a kind of anomaly in our experiment.

For the training samples, they are collected from the real electromagnetic environment by RTL-SDR device. We allocate the center frequency equals 100 MHz, the bandwidth equals 3 MHz and the sampling rate equals 3 M symbols per second. The total number of training samples is 1000.

For the test samples, they can be split into two parts. One part is collected from the real electromagnetic environment. This part is tagged as “0”, which means normal.

The other part is collected from the real electromagnetic environment but processed with a 20 dB and a 30 dB signal-to-noise ratio (SNR), respectively. This part is tagged as "1", which means anomaly. Both parts are composed of 1000 samples.

The experiment is specifically conducted as follows: First of all, take STFTs to both the training data and test data. Next, train conventional one-layer autoencoder network and proposed two-layers autoencoder network separately. For two-layers network, it is trained as follows: first, use the pre-processed input data samples to train the first layer, and then apply the features that first layer has learned to train the second layer. The parameters of the network can be improved according to the reconstruction error using cross-validation method. After that, use the trained network to test the pre-processed test data samples, one-layer autoencoder network and two-layers autoencoder separately, and calculate the reconstruction error. After that, compare the reconstruction error with threshold. If the reconstruction error is beyond threshold, tag the test sample as "1", otherwise, tag the test sample as "0". Finally, compare the label of test data with the output tag, and calculate the accuracy of test samples.

3.6 Experimental Result

We independently run the experiment 10 times and the results we get are shown in Tables 3 and 4. It can be inferred from the experimental results that when the input dataset is processed with a lower SNR which equals 20 dB, both conventional one-layer network and proposed two-layers network can detect the anomalies precisely. But while there is a higher SNR which equals 30 dB processed with input dataset, although conventional one-layer network and proposed two-layers network both have

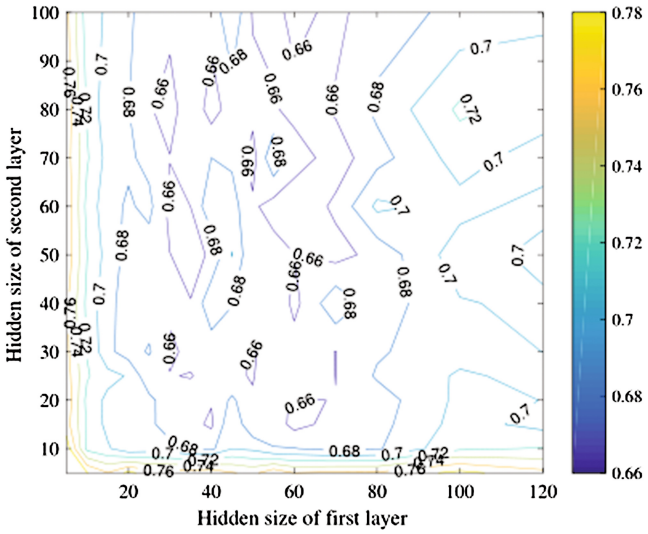


Fig. 2. Contour line of hidden size in neural network. The values of contour lines are magnified 1000 times.

Table 3. Accuracy in different structures when SNR = 20 dB

Network structure	Threshold	Minimum accuracy (%)	Maximum accuracy (%)	Average accuracy (%)
Conventional	4.6×10^{-3}	68.60	100.00	95.52
Proposed	4.6×10^{-3}	**70.55**	100.00	**95.84**

Table 4. Accuracy in different structures when SNR = 30 dB

Network structure	Threshold	Minimum accuracy (%)	Maximum accuracy (%)	Average accuracy (%)
Conventional	9×10^{-4}	56.05	97.40	78.17
Proposed	9×10^{-4}	**61.20**	**97.85**	**80.12**

some wrong predictions to the input. However, the proposed two-layers network reaches a slightly higher accuracy, about 2 %, compared with a conventional one-layer network in this case.

4 Conclusion

We propose a deep autoencoder based method for anomaly detection of spectrum in wireless communication. It relies on the signal spectral features and autoencoder with a deep structure as a one-class classifier. Then we develop a two-layers autoencoder neural network and train it using the time-frequency diagram of input signal which is collected by a RTL-SDR device from the real-time electromagnetic environment. Moreover, we compare the performance of our model with a conventional one-layer autoencoder network, and our proposed model outperforms about 2 % in accuracy when the anomalies are very weak. Future works are intended to use different types of anomalies, such as random impulse and wideband or narrowband unknown signals. Moreover, the parameters and structures of deep auto-encoder neural network remain to be improved.

References

1. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection. ACM Comput. Surv. **41**, 1–58 (2009)
2. Aleskerov, E., Freisleben, B., Cardwatch, R.B.: A neural network based database mining system for credit card fraud detection. In: Proceedings of the IEEE/IAFE 1997: IEEE Computational Intelligence for Financial Engineering (CIFEr), pp. 220–226 (1997)
3. Kumar, V.: Parallel and distributed computing for cybersecurity. IEEE Distrib. Syst. Online, 1 (2005)

4. Bertini, M., Del Bimbo, A., Seidenari, L.: Multi-scale and real-time non-parametric approach for anomaly detection and localization. Comput. Vis. Image Understand. **116**, 320–329 (2012)
5. Khazai, S., Safari, A., Mojaradi, B., Homayouni, S.: An approach for subpixel anomaly detection in hyperspectral images. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. **6**, 769–778 (2013)
6. Quinn, J.A., Sugiyama, M.: A least-squares approach to anomaly detection in static and sequential data. Pattern Recogn. Lett. **40**, 36–40 (2014)
7. Marchi, E., Vesperini, F., Eyben, F., Squartini, S., Schuller, B.: A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks. In: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1996–2000. IEEE (2015)
8. Marchi E, Vesperini F, Weninger F, Eyben F, Squartini S, Schuller B., Non-linear prediction with LSTM recurrent neural networks for acoustic novelty detection. In: 2015 International Joint Conference on Neural Networks (IJCNN), pp. 1–7. IEEE (2015)

A Modified Complex ICA for Blind Source Separation and the Application in Communication Reconnaissance

Zheng Dou(✉), Zi Xiao, Yang Zhao, and Jinyu Wang

Harbin Engineering University, Harbin, China
xiaozi_heu@outlook.com

Abstract. This paper proposes a modified complex ICA algorithm for blind source separation and we apply it in the field of communication reconnaissance. First, Generalized Information Criterion (GIC) and Minimum Description Length (MDL) are used to estimate source signals number. Second, we propose a novel complex independent component analysis (Improved TCMN). The innovation appears in the whitening pre-treatment of TCMN by using the number estimation result of source signals. Finally, this paper gives an application in communication reconnaissance system. Simulation results demonstrate the effectiveness of the new method.

Keywords: Blind source separation · Complex ICA · Communication reconnaissance

1 Introduction

Communication reconnaissance focus on separating mixing signals. Traditional communication reconnaissance signal separation methods are to use different signal characteristics in the transform domain, including the time-domain, the frequency-domain, the code-domain etc. However, those methods require some prior information. Considering the independence and little-to-no priori information of communication signals, BSS technologies are attracted attention to deal with mixed signals [1].

The goal of blind source separation (BSS) is to recover unknown sources based only on their observed mixtures. Applications of BSS can be found in speech signals [2], image processing [3], teleichumedicine [4] etc. Comparing the number of received signals (M) and the number of source signals (N), BSS can be separated into the well-determined and the over-determined BSS with $M \geq N$, and the ill-conditioned case with $M < N$, which is called under-determined BSS (UBSS) [5]. Considering mathematical models of BSS according to different mixing processes, manifestations of BSS cover linear mixing BSS and nonlinear mixing BSS. Several approaches have been proposed by imposing additional assumptions on the sources. A classical method of BSS, independent component analysis (ICA), assumes statistically independent sources [6]. In addition to independence, one of the special properties of sources that can be used for BSS is sparsity, just as [7]. Another one of the special properties that can be

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_43

employed to perform BSS is the nonnegativity of sources and research is presented in papers [8, 9].

With the assumption of independence and combine with complex representation of communication reconnaissance signals, Complex ICA has been introduced to solve separation problems. Complex ICA has been widely studied [10–16]. Transcendental Complex Maximization of Non-Gaussianity (TCMN) [12] is proposed by Novey M and Adali T. This algorithm is proved to be robust for non-Gaussianity. Whitening is an important pre-treatment of ICA, which greatly reduce computation of algorithm. Therefore, the paper proposed an improved TCMN by improving whitening. Simulation results demonstrate the promising performance of the improved TCMN compared to the original TCMN. What's more, this developed algorithm suppresses noise interference effectively. After that, we apply this new method into a justifiable communication reconnaissance separation system. We do several experiments and results show that separation is successful.

2 Research and Application of a Modified Complex ICA

2.1 Source Signals Number Estimation

Covariance of observation data can be used to separate the signals subspace and noise subspace. If source signals number is known, we get a exact boundary of signal subspace and noise subspace.

Observation datas are represented as $\mathbf{X} = \{x_1, x_2, \ldots x_N\}$ and N is observed signals number. $\mathbf{R}_x = \mathrm{E}\{\mathbf{X}\mathbf{X}^{\mathrm{T}}\} = \mathrm{E}\left\{(\mathbf{X} - \mathrm{E}\{\mathbf{X}\})(\mathbf{X} - \mathrm{E}\{\mathbf{X}\})^{\mathrm{H}}\right\}$ is the covariance matrix of $\mathbf{X}$. We use "k" $(1 \le k \le N)$ to represent the rank of $\mathbf{R}_x$ and use $\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_k$ to represent eigenvalues. "N_s" is the number of signal sampling points. We display two practical estimation criterias based on information theory. They are Generalized Information Criterion (GIC) [17] and Minimum Description Length (MDL) [18, 19].

$$GIC(k) = -2N_s(N-k)\log\rho(k) + \rho k(2N-k) \tag{1}$$

$$MDL(k) = -N_s(N-k)\log\rho(k) + \frac{1}{2}k(2N-k)\log N_s \tag{2}$$

In which, ρ from 2 to 6, all the $\rho(k)$ are same, as shown below.

$$\rho(k) = \frac{(\lambda_{k+1}\lambda_{k+2}\ldots\lambda_N)^{\frac{1}{(N-k)}}}{\frac{1}{(N-k)}(\lambda_{k+1} + \lambda_{k+2} + \ldots + \lambda_N)} \tag{3}$$

Choosing one of this two criterions, we can estimate source signals number $\hat{M} = k$.

2.2 Whitening Improvement

Eigenvalues of signals subspace which can be used to compute whitening matrix of ICA algorithm. Classic whitening does not make full use of eigenvalues of noise subspace. Exploiting eigenvalues of noise subspace to estimate noise variance which can be used to correct the eigenvalues of signal subspace.

Eigenvalue decomposition of covariance matrix is shown below.

$$\mathbf{R}_x = \mathbf{U}\Sigma\mathbf{U}^H + \sigma^2\mathbf{I} = \mathbf{U}(\Sigma + \sigma^2\mathbf{I})\mathbf{U}^H = \mathbf{U}\Lambda\mathbf{U}^H \tag{4}$$

$$\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_N) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2) \tag{5}$$

$\sigma_i^2, (i = 1, 2, \ldots, M)$, is the sequence of eigenvalues of signal subspace, and unitary matrix is expressed as $\mathbf{U} = [u_1, u_2, \ldots, u_M, u_{M+1}, \ldots, u_N]$.

Eigenvalues and eigenvectors of signals subspace are

$$\mathbf{\Lambda}_s = diag(\lambda_1, \lambda_2, \ldots, \lambda_M) = diag(\sigma_1^2, \sigma_2^2, \ldots, \sigma_M^2) \tag{6}$$

$$\mathbf{U}_s = [u_1, u_2, \ldots, u_M] \tag{7}$$

Estimating noise variance

$$\hat{\sigma}^2 = \frac{\lambda_{M+1} + \ldots + \lambda_N}{N - M} \tag{8}$$

Regularizing $\hat{\mathbf{\Lambda}}_s$ by using $\hat{\sigma}^2$,

$$\hat{\mathbf{\Lambda}}_s = diag(\sigma_1^2 - \hat{\sigma}^2, \sigma_2^2 - \hat{\sigma}^2, \ldots, \sigma_M^2 - \hat{\sigma}^2) \tag{9}$$

Finally, the improved whitening matrix is presented as

$$\mathbf{V}_s = \hat{\mathbf{\Lambda}}_s^{-1/2}\mathbf{U}_s^{\mathrm{H}} \tag{10}$$

2.3 Complex ICA

- **The Modified TCMN**

  Objective function of TCMN

$$J(\mathbf{w}) = E\left\{\left|G\left(\mathbf{w}^{\mathrm{H}}\mathbf{z}\right)\right|^2\right\} \tag{11}$$

Main nonlinear function $G(y)$ and first derivative $g(y)$ of TCMN algorithm are shown in formulas (12) and (13).

$$G_1(y) = a\ \sinh(y),\ \ G_2(y) = \cos(y),\ \ G_3(y) = y^2 \tag{12}$$

$$g_1(y) = \frac{1}{\sqrt{y^2+1}},\ \ g_2(y) = -\sin(y),\ \ g_3(y) = 2y \tag{13}$$

With the modification of whitening pre-treatment, we propose the modified TCMN algorithm. Algorithm steps is shown in Table 1.

- **Simulation**

Giving two source signals 4QAM. First one sampling rate is $f_s = 500\,\text{kHz}$, carrier frequency $f_c = 0\,\text{kHz}$, information rate $f_b = 20\,\text{kbps}$ and the other one sampling rate is $f_s = 500\,\text{kHz}$, carrier frequency $f_c = 0\,\text{kHz}$, information rate $f_b = 10\,\text{kbps}$. Simulation point number is 5,000.

We use average crosstalk error PI to evaluate separation performance. The smaller the value, the better the performance.

$$PI = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left\{\left(\sum_{k=1}^{n}\frac{|g_{ik}|}{\max_j|g_{ij}|}-1\right)+\left(\sum_{k=1}^{n}\frac{|g_{ki}|}{\max_j|g_{ji}|}-1\right)\right\} \tag{14}$$

Here, $\mathbf{G} = \mathbf{WA}$ is global matrix. When our algorithm can be used to separate source signals, $\mathbf{G}$ tends to be a unit matrix. In fact, when PI is approximately equal to 10^{-2}, the separation performance is great. As shown in Fig. 1, separation performance of Improved TCMN algorithm is better than TCMN, especially in a low SNR.

Table 1. Improved TCMN algorithm steps

1. Improved whitening pretreatment for observational data $\mathbf{x}$, $\mathbf{z} = \mathbf{Vx}$, initializing p = 1 in the same time.
2. Initializing matrix $\mathbf{w}_p$ whose normal number is 1.
3. Update and orthogonalize $\mathbf{w}_p$ $\mathbf{w}_p \leftarrow -E\{G^*(\mathbf{y})g(\mathbf{y})\mathbf{x}\} + E(g(\mathbf{y})g^*(\mathbf{y}))\mathbf{w}_p$ $+ E\{\mathbf{x}\mathbf{x}^T\}E\{G^*(\mathbf{y})g'(\mathbf{y})\}\mathbf{w}_p^*$ $\mathbf{w}_p = \mathbf{w}_p - \sum_{j=1}^{p-1}\mathbf{w}_j\mathbf{w}_j^{\mathrm{H}}\mathbf{w}_p,\ \mathbf{w}_p = \frac{\mathbf{w}_p}{\|\mathbf{w}_p\|}$
4. If $\mathbf{w}_p$ not convergence, return step 3.
5. $p = p+1$, if $p \le n$ (n is the number of source signals), return step 2.
6. $\mathbf{y} = \mathbf{W}^{\mathrm{H}}\mathbf{x}$ separates source signals, separation matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_n]^{\mathrm{T}}$

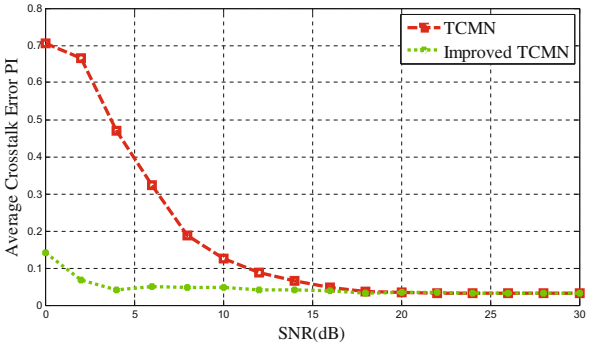


Fig. 1. TCMN and improved TCMN separation performance comparison

3 Application of the Modified TCMN

The modified TCMN algorithm is appropriate for communication reconnaissance area. For example, Received signal of the pilot often subject to interference from other signals and noise during the flight. Exploiting the improved algorithm to separating signals and getting what we want.

- **Simulation Experiment**

For validating the effectiveness of our new algorithm in communication reconnaissance area. We do experiment that two FM source signals and two received signals. Signals are overlap whether in the time-domain or frequency-domain. Using our

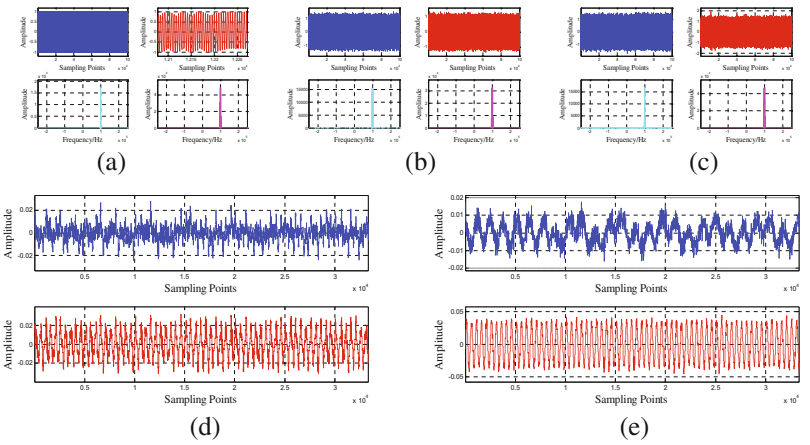


Fig. 2. (a) is the time-domain and frequency-domain diagrams of source signals. (b) is the time-domain and frequency-domain diagrams of mixing signals. (c) is the time-domain and frequency-domain diagrams of separated signals. (d) is the time-domain waveform of demodulation without separation. (e) is the time-domain waveform of demodulation with separation.

algorithm to estimate mixing matrix and to separate source signals. The modulated signal of the first FM is a speech signal, sampling rate of modulation signal is $f_b = 16\,\text{kHz}$, modulation index $m_f = 6$, sampling rate $f_s = 500\,\text{kHz}$, and carrier frequency $f_c = 100\,\text{kHz}$. The second FM's modulation signal is a sine signal with $f_b = 1\,\text{kHz}$, modulation index $m_f = 6$, sampling rate $f_s = 500\,\text{kHz}$, and carrier frequency $f_c = 100\,\text{kHz}$. Simulation sampling number is 5,000. As shown in Fig. 2, the improved TCMN algorithm can be use to separate communication reconnaissance signals feasibility.

4 Conclusion

This research aims to propose a solution method for separating communication reconnaissance signals when source signals number is unknown in complicated electromagnetic environment. In this paper, We improve whitening by using the result of source signals number estimation that separating signal subspace and noise subspace accurately and rectify eigenvalues of signal subspace by exploiting noise variance to inhibit noise. And then, we proposed the modified TCMN algorithm by exploiting improved whitening. We verify the superiority of the improved TCMN algorithm compared with TCMN algorithm and the feasibility of the improved TCMN for a separation application via simulation in communication reconnaissance area.

References

1. Fu, W., et al.: Blind source separation algorithm for communication complex signals in communication reconnaissance (2009)
2. Li, S., et al.: BSS algorithm based on fully connected recurrent neural network and the application in separation of speech signals, 1–3 (2012)
3. Karray, E., Loghmari, M.A., Naceur, M.S.: Blind source separation of hyperspectral images in DCT-domain, 381–388 (2010)
4. Sahroni, A., Setiawan, H., Marfianti, E.: Performance of blind source separation (BSS) techniques for mixed source signals of EEG, ECG, and voice signal, 213–217 (2014)
5. Xie, S., et al.: Time-frequency approach to underdetermined blind source separation **23**, 306–311 (2012)
6. Comon, P., Jutten, C.: Handbook of bind source separation: independent component analysis and applications (2009)
7. Xie, S., et al.: Time-frequency approach to underdetermined blind source separation, 306–316 (2012)
8. Yang, Z., Rong, Y.: A convex geometry-based blind source separation method for separating nonnegative sources **26**, 1635–1644 (2015)
9. Guan, N., Tao, D., Luo, Z., Yuan, B.: Online nonnegative matrix factorization with robust stochastic approximation (23), 1087–1099 (2012)
10. Bingham, E., Hyvarnen, A.: A fast fixed-point algorithm for independent component analysis of complex valued signals **10**, 1–8 (2000)
11. Novey, M., Adali, T.: On extending the complex Fast-ICA algorithm noncircular sources **56**, 2148–2154 (2008)

12. Novey, M., Adali, T.: Complex ICA by negentropy maximization, 596–609 (2008)
13. Rajapakse, J.C., Chen, W.: Complex ICA-R, 1–8 (2010)
14. Li, X.-L., Adali, T.: Complex independent component analysis by entropy bound minimization, 1417–1430 (2010)
15. Yang, Z., Kassam, S.A.: Optimum nonlinearity and approximation in complex Fast-ICA, 1–6 (2012)
16. Fu, G.-S., Phlypo, R., Anderson, M.: Complex independent component analysis using three types of diversity, 794–805 (2015)
17. Petre, S., Yngve, S.: A review of information criterion rules (2004)
18. Schwartz, G.: Estimating the dimension of a mode, 243–258 (1978)
19. Xiao, M., Wei, P., Tai, H.: Estimation of the number of sources based on hypothesis testing, 465–477 (2012)

Sensor Coverage Problem in Sparse MANET Environments

JongBeom Lim[1], HeonChang Yu[2], and JoonMin Gil[3(✉)]

[1] IT Convergence Education Center, Dongguk University, Seoul, Korea
jblim@dongguk.edu
[2] Department of Computer Science and Engineering, Korea University, Seoul, Korea
yuhc@korea.ac.kr
[3] School of Information Technology, Catholic University of Daegu, Gyeongsan, Korea
jmgil@cu.ac.kr

Abstract. In this paper, we define the problem of sensor coverage in sparse mobile ad hoc networks. Previously, the nodes are assumed static or the number of nodes is large enough to cover the area for the coverage problem in wireless sensor networks. However, in sparse mobile ad hoc network environments, the semantics of the coverage problem differ in that the nodes are free to move in the area and the distance between the nodes should be long enough in order not to overlap the nodes' coverage to maximize the total coverage area in the network. We formulate the sensor coverage problem in sparse mobile ad hoc network environments.

Keywords: Sensor coverage · Mobile ad hoc network · Wireless sensor network · Machine to machine

1 Introduction

Wireless sensor networks (WSNs), sometimes called wireless sensor and actuator networks (WSANs), are spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, pressure, etc. and to forward their data via the network [1]. Meanwhile, mobile ad hoc networks (MANETs) have been one of the most innovative and challenging wireless networking paradigms in the field [2, 3]. Hence, MANET immediately gained momentum and produced tremendous research efforts in the mobile network community [4–6].

By taking advantage of WSN and MANET, there are many applications that projected to have a significant impact into our daily lives such as battlefield surveillance, environmental monitoring, industrial diagnostics, and precision agriculture [7, 8].

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2055463 and NRF-2015R1D1A1A01061373).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_44

Among various performance metrics for sensor networks, coverage is one of the most important ones that reflect how well a sensor field is monitored [9, 10].

The coverage problem is a well-known problem in wireless sensor networks. To maximize the coverage area, the nodes in the network should cooperate with each other. Because the sensor nodes have limited capabilities including battery, computing power, and coverage, each node proactively schedules itself to be in a state either sleep or wake-up with the assumption of the dense network. Covering the area of the sensor networks with minimum nodes is considered as an NP-hard problem and various solutions have been proposed [11–13].

On the other hand, when few nodes exist in the sensor network, it is necessary to reconsider the sensor coverage problem. With this assumption, when the nodes in the network are situated at the near location, the coverage problem can be considered the same as in dense network environments, covering a limited area. Each node in a MANET, however, is free to move independently in any direction and is capable of changing its links to other nodes frequently. In this regard, to maximize the coverage, the nodes in the MANET should move at other locations in order not to overlap the sensor coverage of other nodes.

In this case, the coverage problem is different from the case that assumes the network is dense. More specifically, the scheduling such as sleep or wake-up is dismissed, while determining nodes' location and movement is more important to maximize the coverage with the nodes in a MANET.

For instance, suppose that there are 50 nodes in the network. When all the 50 nodes are located at the same location in the area, sensing information of 49 nodes is useless since this information is duplicated.

Because there is no central authority or coordinator of the network and each node has knowledge of local information, maximizing the coverage area in sparse MANET environments is non-trivial task. In such a circumstance, determining moving directions and distance of a node can be done only by its own decision, and this local decision is a key factor that affects the global coverage area in the network. How to realize such local decision that maximizes the coverage area is at the core of our approach.

2 System Model and Problem Definition

2.1 System Model

The network consists of a collection of n nodes, $node_1$, $node_2$, $node_3$, $\cdots$, $node_n$, where each node is connected to one or several sensors and has network capable components: a radio transceiver with an antenna, a microcontroller, and an electronic circuit for interfacing with various sensors. Size and cost constraints on sensor nodes result in corresponding constraints on resources such as energy, memory, computational speed, and communications bandwidth. For an energy source, we assume that each node has an embedded form of energy harvesting (i.e., solar panels). The coverage of the sensors is equal to each other.

The topology of the network is an advanced multi-hop wireless mesh network, where each node maintains a subset of the nodes in the network. The propagation and aggregation techniques between the hops of the network can be routing. Each node has no global information of the system and the nodes communicate solely by passing messages. Message sending and receiving are done in an asynchronous way and all decisions are made in rounds. In other words, sensor nodes do not follow a common clock and each node makes its own decision about movement for a period of time based on its individual clock.

Messages are delivered reliably with finite but arbitrary time delay. The network can be described as a directed graph, in which vertices represent the nodes and edges represent unidirectional communication channels. With an assumption of MANET, each node is free to move independently in any direction within a predefined area. Sensing and communication are modeled as unit disk graphs (UDGs), with the corresponding sensing radius (SR) and communication radius (CR). In the UDG model, a sensor is able to monitor location of an event if and only if the location is within the SR of the sensor. Two nodes are communication neighbors if they are within CR distance from each other and it is assumed that SR $\ll$ CR.

2.2 Problem Definition

One of the fundamental issues in WSN and MANET is the sensor coverage problem. Sensor coverage is to deploy a set of sensor nodes in an area of interest for monitoring and tracking. Sensor nodes are normally densely deployed in the network. To prolong the network lifetime, sensors should sleep as much as possible. In this case, they should wake up only when they are really needed. However, when sensor nodes are sparsely deployed, the problem of sensor coverage is different. We consider a sparse circumstance, where $\sum_{i=1}^{n} SR_{node_i} \ll Total\,Coverage\,Area$.

In such a case, the sensor nodes in the system should be well deployed not to overlap their SRs to maximize the coverage area, while the network should not be partitioned (i.e., a sensor node's CR should be overlap with another node's CR). This makes solving the sensor coverage problem in sparse environments more difficult. To solve this problem, the following conditions should be satisfied: (1) Minimum overlap of SR, and (2) No network partition.

- min $(\forall i, j\ \exists node_i, node_j\ [node_i \neq node_j : i, j \in \{1 \ldots n\} \Rightarrow SR(node_i) \cap SR(node_j)])$ **(Minimum overlap of SR)**
- $\forall i, j\ \neg\exists node_i, node_j\ [node_i \neq node_j : i, j \in \{1 \ldots n\} \Rightarrow CR(node_i)\ \emptyset\ CR(node_j)]$ **(No network partition)**

3 Our Approach

In this section, we describe our approach for maximizing sensor coverage while satisfying the no network partition requirement. There are three steps: (1) local decision for neighbor detection, (2) movement decision when it finds a near neighbor, and

(3) actual node movement in order not to overlap the coverage area with the target node. When Node A detects Node B, whose coverage area is overlap with its SR, it calculates the distance between Node A and Node B. Then, Node A makes a decision of the movement direction and distance.

What makes solving the sensor nodes movement in sparse environments particularly difficult is that each node has no global information of the network and has to make a decision based solely on its local information. Because each node maintains partial node information, it is necessary to realize a mechanism that makes a local decision possible.

For local decisions, there are two different kinds of threads in each node: active and passive. At each round, an active thread selects a neighbor at random and sends a message. The active thread then waits for the message from the receiver. Upon receiving the message from the neighbor, the active thread updates local information with the received message and its own information. A passive thread waits for messages sent by active threads and replies to the senders. Afterwards, the passive thread updates its local information with the received message from the sender accordingly. The discoverNeighbor() function returns a neighbor identifier and access point (e.g., ip address) within its CR range and each node maintain its partial node list as it encounters.

Because of the inherited nature of MANET, that is, each node should perform self-organization and self-configuring processes continuously and maintain the neighbor node information at each round based on the current location. The main purpose of the nearest node discovery is to make the algorithm efficient before each node performs the movement process. Because maximizing the coverage area is largely determined by the nearest node in the network, the nearest node discovery step should be performed with current location information to minimize the overlapping area between the nodes.

The output of the nearest node discovery algorithm is the nearest node for $Node_i$. Obviously, the nearest node discovery algorithm is performed at each round because the nodes' locations are continuously changing. For all $Node_j$ in partial node list, it calculates the distance with $Node_j$. Because the nodes are distributed in 2-dimensional geometry, it first calculates difference of x and y locations. Afterwards, it performs *dist ()* function that returns the square root of the sum of squares of x and y. Then, $Node_i$ selects the nearest node among nodes in $NodeList_i$ and assigns the nearest node.

The simplest way to solve the movement problem is to move at random direction with random distance. This approach, however, does not satisfy minimum overlap of SR and no network partition requirements. Hence, we take another approach by considering SR and CR so that $SR(node_i) \cap SR(node_j)$ is minimized while $\neg (CR(node_i)\ \emptyset\ CR(node_j))$ after the movement process is performed.

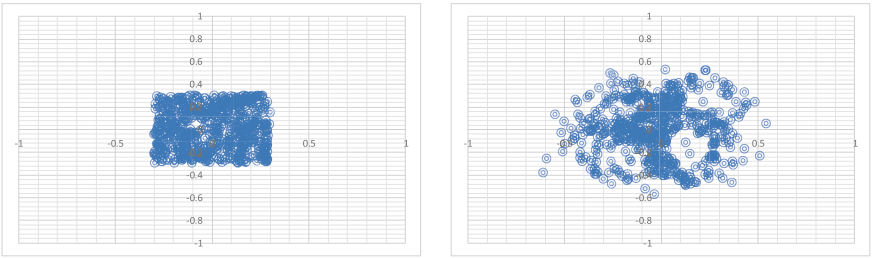
There are four steps for the node movement algorithm. First, it checks whether the distance with $Node_j$ is below $Threshold_{CR}$. When this is the case, it is necessary to move to another location in order not to overlap the coverage area with $Node_j$. Otherwise, it stays with current location to save energy consumption. Second, if it determined to move the location, it decides the direction for movement such that the CR overlapping area is minimized by moving minimum distance. Therefore, $Node_i$ should move in opposite direction with $Node_j$. Third, it determines the moving distance

based on $diff_x$ and $diff_y$. Note that *ScaleFactor* is used for calculating moving distance and this parameter can be configured as a system-wide parameter according to SR and CR. Lastly, based on the determined x and y coordinates, it performs the movement process and change its location. At the next round, the whole procedures (local decision, nearest node discovery, and movement algorithm) will be repeated.

4 Performance Evaluation

In this section, we present experimental results that demonstrate the performance of our presented algorithms. There are 500 nodes in the 2-dimentional MANET relative space from –1 to 1. We consider the worst-case scenario such that all the 500 nodes are located at [0, 0] at the beginning. For experimental parameters, $Threshold_{CR}$ is set to 0.1 and *ScaleFactor* is set to 0.2. For each round, each node performs nearest node discovery and movement algorithms

Figure 1 shows the experimental results after 1 round and 20 rounds. As shown in Fig. 1(a), the nodes are distributed covering about [± 0.3, ± 0.3]. After 20 rounds (Fig. 1(b)), the nodes are further distributed in the space. Note that the results will be different when the different values of the experimental parameters (e.g., $Threshold_{CR}$, *ScaleFactor*) are chosen. In this regards, it is also important to select the optimal values of the parameters with careful consideration for the number of nodes, SR, CR, etc. Because we are focusing on validating the algorithms, we leave this as future work.



(a) After 1 round (b) After 20 rounds

Fig. 1. Experimental results.

5 Conclusions

In this paper, we defined the sensor coverage problem in sparse MANET environments and proposed its solution using one-to-one communication primitives, which is more efficient in comparison to broadcast primitives. Our proposed approach can be deployed in wide variety of networks, and system wide parameters such as $Threshold_{CR}$ and *ScaleFactor* can be tuned into a particular environment for one's purpose. Even in the worst-case scenario, where all the nodes are situated in the same coordinates, our algorithms perform well in terms of the total coverage area of the nodes

while satisfying the CR ranges for communications. Future work includes employing a variant of a genetic algorithm to further enhance the performance and distribute the nodes more evenly.

References

1. Fadel, E., Gungor, V.C., Nassef, L., Akkari, N., Malik, M.G.A., Almasri, S., Akyildiz, I.F.: A survey on wireless sensor networks for smart grid. Comput. Commun. **71**, 22–33 (2015)
2. Conti, M., Giordano, S.: Mobile ad hoc networking: milestones, challenges, and new research directions. IEEE Commun. Mag. **52**, 85–96 (2014)
3. Conti, M., Giordano, S.: Multihop ad hoc networking: the evolutionary path. In: Mobile Ad Hoc Networking, pp. 1–33. Wiley, New York (2013)
4. Kim, D., Kim, J.-h., Moon, C., Choi, J., Yeom, I.: Efficient content delivery in mobile ad-hoc networks using CCN. In: Ad Hoc Networks, vol. 36, Part 1, pp. 81–99 (2016)
5. Zhang, X.M., Wang, E.B., Xia, J.J., Sung, D.K.: A neighbor coverage-based probabilistic rebroadcast for reducing routing overhead in mobile ad hoc networks. IEEE Trans. Mob. Comput. **12**, 424–433 (2013)
6. Sun, W., Yang, Z., Zhang, X., Liu, Y.: Energy-efficient neighbor discovery in mobile ad hoc and wireless sensor networks: a survey. IEEE Commun. Surv. Tutorials **16**, 1448–1459 (2014)
7. Bellavista, P., Cardone, G., Corradi, A., Foschini, L.: Convergence of MANET and WSN in IoT urban scenarios. IEEE Sens. J. **13**, 3558–3567 (2013)
8. Conti, M., Boldrini, C., Kanhere, S.S., Mingozzi, E., Pagani, E., Ruiz, P.M., Younis, M.: From MANET to people-centric networking: milestones and open research challenges. Comput. Commun. **71**, 1–21 (2015)
9. Mini, S., Udgata, S.K., Sabat, S.L.: Sensor deployment and scheduling for target coverage problem in wireless sensor networks. IEEE Sens. J. **14**, 636–644 (2014)
10. Liu, B., Dousse, O., Nain, P., Towsley, D.: Dynamic coverage of mobile sensor networks. IEEE Trans. Parallel Distrib. Syst. **24**, 301–311 (2013)
11. Luo, W., Wang, J., Guo, J., Chen, J.: Parameterized complexity of max-lifetime target coverage in wireless sensor networks. Theor. Comput. Sci. **518**, 32–41 (2014)
12. Liao, Z., Wang, J., Zhang, S., Cao, J., Min, G.: Minimizing movement for target coverage and network connectivity in mobile sensor networks. IEEE Trans. Parallel Distrib. Syst. **26**, 1971–1983 (2015)
13. Lu, Z., Li, W.W., Pan, M.: Maximum lifetime scheduling for target coverage and data collection in wireless sensor networks. IEEE Trans. Veh. Technol. **64**, 714–727 (2015)

Design of Jitter Buffer Control Algorithm for Guaranteeing the Medical Information Data Transmission Quality in Wireless Network Environment

Dong-Wan Joe[1], Jae-Sung Shim[1], Yong-Wan Ju[2], and Seok-Cheon Park[3(✉)]

[1] Department of IT Convergence Engineering, Gachon University, Gyeonggi-do, South Korea
{jfdwe,llsjs28}@naver.com
[2] Correspondence Center of Korea Internet and Security Agency, Seoul, South Korea
ywju@naver.com
[3] Department of Computer Engineering, Gachon University, Gyeonggi-do, South Korea
scpark@gachon.ac.kr

Abstract. Telemedicine that reduces the time and spatial restrictions on healthcare is expected to be broadly activated due to the recent development of ICT convergence technologies. The VoIP technology is being used for data transfer in the telemedicine environment. In the IP network that uses the packet exchange method, however, problems of delay, jitter, and packet loss, among others, are occurring. Therefore, the condition of the VoIP network was predicted in this study using a network condition prediction algorithm. A VoIP-based dynamic jitter buffer control algorithm that addresses transfer quality problems through dynamic jitter buffer control was proposed.

Keywords: VoIP · Jitter buffer · Prediction of the network condition

1 Introduction

Telemedicine that reduces the time and spatial restrictions on healthcare is expected to be broadly activated due to the recent development of ICT convergence technologies [1].

Telemedicine uses VoIP, which is inexpensive in terms of the communication cost, and is easy to use.

In the current IP network that uses the packet exchange method, however, problems of delay, packet loss, and packet sequence switch, among others, are occurring [2].

In the wireless network environment, the problems of image and voice interruption, echoing sound, and image and voice discrepancy occur due to the transfer delay of VoIP, or packet loss.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_45

To address these problems, the network condition is optimized through jitter buffer control, but it is difficult for the existing fixed jitter buffer to respond to the variably changing wireless network environment.

Therefore, the condition of the wireless network environment was predicted in this study through a network condition prediction algorithm, and a VoIP-based dynamic jitter buffer control algorithm that can guarantee the transfer quality through dynamic buffer control was designed and evaluated.

In Sect. 2 of this paper, the VoIP that is used in telemedicine and an algorithm for predicting the network condition are analyzed. In Sect. 3, the design of a dynamic jitter buffer control algorithm for improving the VoIP-based medical information transfer quality is presented. Finally, in Sect. 4, the study conclusions are drawn.

2 Related Studies

2.1 VoIP(Voice Over Internet Protocol)

VoIP is the technology for exchanging voices or data through the Internet. VoIP uses the IP network instead of the previous public switched telephone network, and it sends voice information in the form of discontinuous digital packets [3].

VoIP combines the characteristics of the data network that tries to maintain a continuous connection and those of the voice communication that occurs sporadically.

It converts the analog signals of the data network voice to digital signals, compresses the signals, converts the signals into IP packets, and transfers the packets [4]. Figure 1 shows the communication process of VoIP.

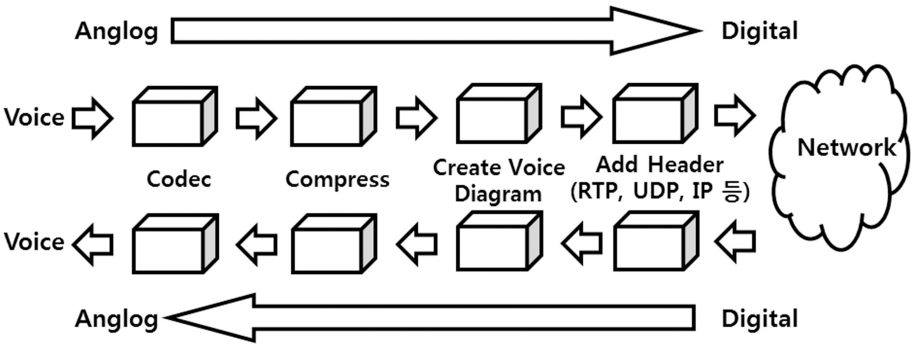


Fig. 1. VoIP communication process

2.2 Jitter Buffer

Jitter buffer is an algorithm for maintaining the optimal balance on packet delay or packet loss.

If the jitter buffer for removing jitter is too large, the end-to-end delay is increased as much because the packets stay longer in the buffer [5].

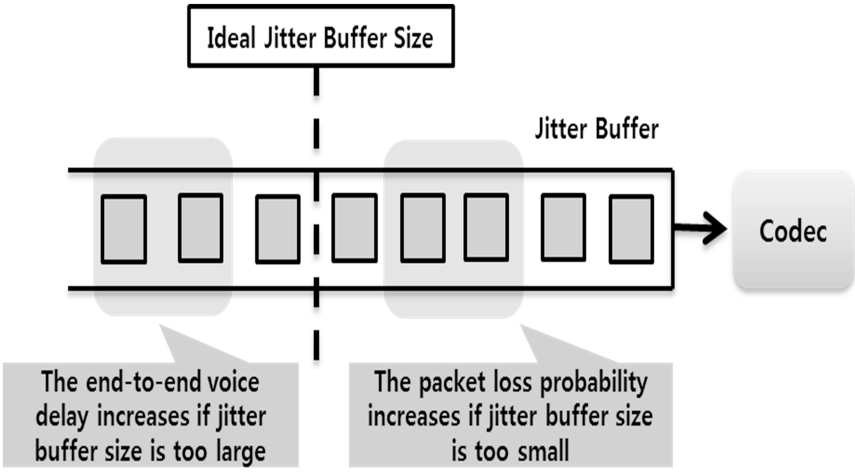


Fig. 2. Jitter buffer structure

If the jitter buffer is too small, data loss may occur. Thus, the size of the jitter buffer must be appropriately determined after checking the network and terminal condition through a jitter buffer algorithm [6]. Figure 2 shows the structure of the jitter buffer.

2.3 Prediction of the Network Condition

Predicting the network traffic requires periodic models such as the continuous time analysis model, the neural network method, and the wavelet method.

It requires a long period of time and involves complicated calculations. For real-time data exchanges including VoIP, however, short-term predictions for such cases as network traffic monitoring, congestion control, and intrusion detection are required.

Least mean square is a statistical algorithm that can accept the requirements of short-term predictions.

It can guarantee less calculations and reliability. Especially, it can be applied in real time [7, 8].

3 Design of the Dynamic Jitter Buffer Control Algorithm

3.1 Algorithm Overview

The dynamic jitter buffer control algorithm proposed in this study collects the time delay data of the received packets and predicts the network condition with the collected time delay data using a prediction algorithm.

Also, it appropriately controls the size of the jitter buffer based on the predicted network condition information, and improves the transfer quality by addressing the transfer delay in the VoIP network, or the packet loss problems. Figure 3 is the overall diagram of the algorithm proposed in this study.

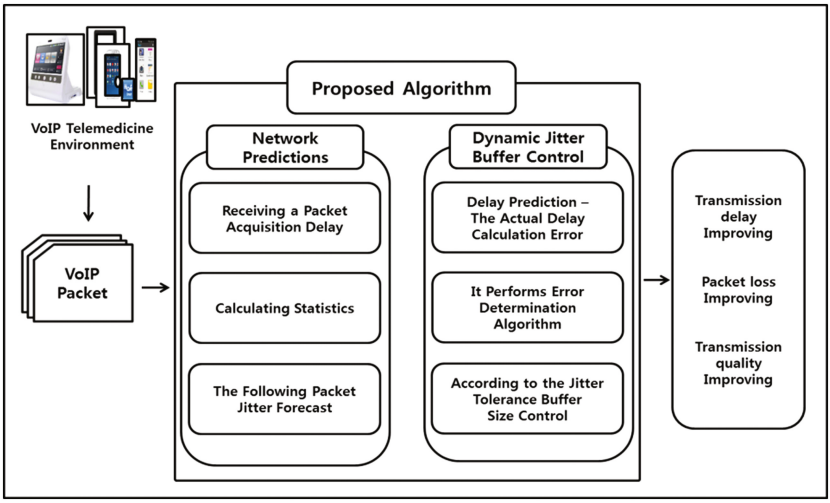


Fig. 3. Algorithm overview

3.2 Dynamic Jitter Buffer Control Algorithm

The dynamic jitter buffer algorithm proposed in this study controls the size of the jitter buffer by reading the header data, such as the time stamps of the packets, and reflecting the network delay condition and jitter.

Also, it has the advantage of finding the optimal buffer size compared to the existing jitter buffer because it calculates the size of the jitter buffer by predicting the network condition.

$$\mathrm{T_{i,min}} = \frac{\mathrm{Min}}{j}\left\{\mathrm{T}_{j,net}\right\} = 1, \ldots, \mathrm{i}-1 \tag{1}$$

$$\mathrm{B_{i-1}} = \mathrm{T}_{i-1} - \mathrm{T}_{i,\mathrm{min}},\ \text{for}\,\mathrm{i}\ > 1 \tag{2}$$

$$\begin{aligned} B_i' &= \mathrm{B}_{i-1}(1+\mathrm{f}) \\ \mathrm{B}_i &= B_i' + \mathrm{T}_{i,\mathrm{min}} + \mathrm{T}_{i+1,\mathrm{min}} \end{aligned} \tag{3}$$

$$J_{e_{i+1}} = avgJ_i + \beta_i \times \mathrm{var}J_i \tag{4}$$

$$optB_i = B_i' + \mathrm{T}_{i,\mathrm{min}} - \mathrm{T}_{i+1,\mathrm{min}} + J_{e_{i+1}} \tag{5}$$

Equations 1, 2, 3, 4, and 5 are used in the dynamic jitter buffer algorithm proposed in this study.

$\mathrm{T}_{i,\mathrm{min}}$ in Eq. 1 represents the shortest delay among the network delays of the packets received up to i-th meaning the fixed delay.

In Eq. 2, the value obtained by subtracting the fixed delay from the network delay of the i–1th packet is determined as the size of the jitter buffer, B_{i-1}, which means the network variable delay.

In Eq. 3, the f factor is added, and the initial value of β_i, B'_i is derived. The size of the jitter buffer, B_i, is derived by adding the variable delay to B'_i.

In Eq. 4, the network jitter, $J_{e_{i+1}}$, of the packet to be received next is judged using the average and variance of the network jitter as well as the variance weight.

In Eq. 5, a more optimized buffer size, $optB_i$, is obtained by reflecting the predicted jitter value of the next packet, $J_{e_{i+1}}$, calculated using the network condition prediction algorithm, in the buffer size, β_i, of Eq. 4, which reflects the delay condition.

The error between the predicted jitter value by $optB_i$ and the actual jitter value is obtained.

If there is no error, the buffer size is fixed. If there is an error, the error size is calculated.

An error bigger than 0 means that the actual network condition traffic is smaller than the predicted condition; thus, the size of the jitter buffer for the next packet is reduced.

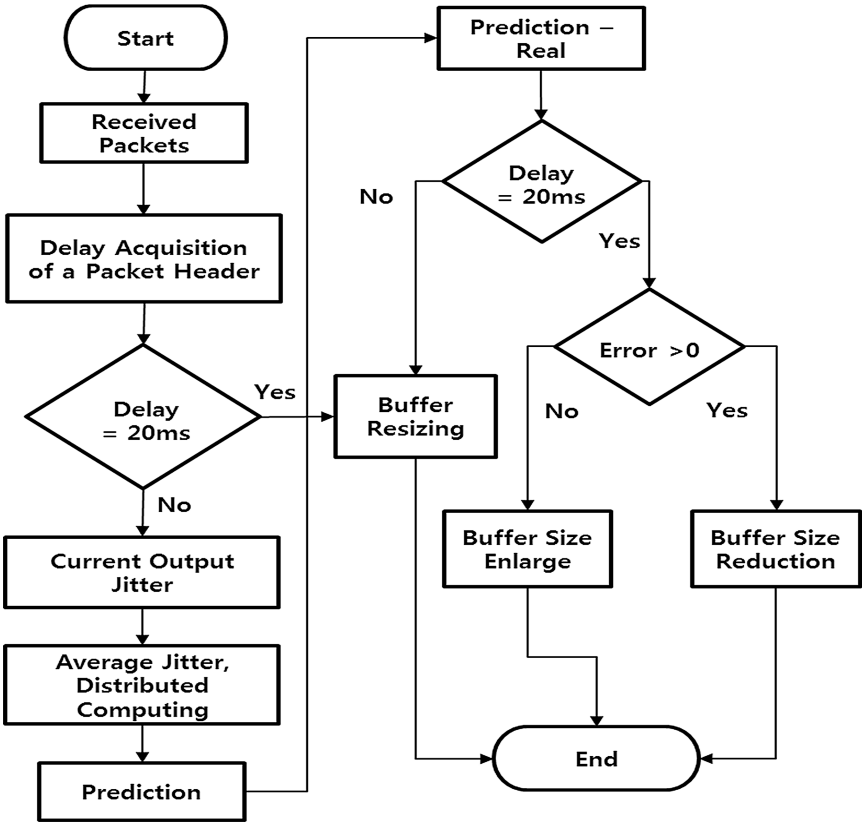


Fig. 4. Dynamic jitter buffer control algorithm

An error bigger than 0 means that the network traffic condition is not better than the predicted condition; thus, the size of the buffer increases. Figure 4 shows the VoIP-based dynamic jitter buffer control algorithm.

4 Conclusion

The telemedicine business is growing due to the recent development of medical and wireless network technologies. It not only allows the transmission and receipt of biological data but also provides video medical treatment between doctors and patients.

Due to such changes in remote medical treatment services, the proportion of multimedia data, including voice, image, and video, is increasing, which leads more medical equipment developers for remote medical treatment to use the VoIP technology, which is inexpensive in terms of the communication cost and makes it easy to transfer multimedia data.

The transfer quality of VoIP must be guaranteed for the remote medical treatment environment, which requires high reliability, but the problems of delay, packet loss, and packet sequence switch, among others, are occurring in the IP network that uses the packet exchange method.

Therefore, a VoIP-based jitter buffer control algorithm for guaranteeing the medical information data transfer quality that addresses transfer quality problems through the prediction of the VoIP network condition and dynamic jitter buffer control was designed in this study.

References

1. Kim, H.-J., Huh, J.-S.: The problems of pilot projection for telemedicine and law of medicine. Korean J. Med. Law **23**(1), 7–20 (2015)
2. Kim, G.: (A) design for anonymous authentication protocol for user information protection in u-health care environment, Ajou University master's thesis (2014)
3. Choi, D.: Effects of communication environment on VoIP capacity using WiFi. Int. J. Inf. Commun. Eng. **19**(6), 1327–1332 (2015)
4. Choei, S.: Study for performance improvement of jitter buffer algorithm applied prediction model, Kongju National University Ph.D. thesis (2014)
5. Kim, H.: (A) study on improvement of VoIP network quality by optimizing jitter buffer size, Dankook University master's thesis (2013)
6. Choi, S.-H., Park, J.-M., Seo, C.-H.: The analysis of event-based jitter buffer algorithm. J. Korea Inst. Inf. Secur. Crypt. **23**(5), 867–871 (2013)
7. Kim, H.: Network delay prediction based robust adaptive playout algorithm, Catholic University of Korea master's thesis (2008)
8. Seo, K.-D., Lee, J.-H., Kim, H.-G.: Network jitter estimation algorithm for robust VoIP system in vehicle environment. J. Korea Inst. Intell. Transp. Syst. **10**(4), 93–99 (2011)

Student's-t Mixture Model Based Excepted Patch Log Likelihood Method for Image Denoising

J.W. Zhang[1(✉)], J. Liu[1], Y.H. Zheng[2], and J. Wang[2]

[1] College of Math and Statistic, Nanjing University of Information Science and Technology, Nanjing 210044, China
zhangjw@nuist.edu.cn
[2] CICAEET, College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract. Recently, patch priors based image denoising method has received much attention in recent years. Expected patch log likelihood (EPLL) is a popular method with the patch priors for image denoising, which achieves image noise removal using the Gaussian mixture priors learned by the Gaussian mixture model (GMM). In this paper, with observation that the student's-t distribution has a heavy tail and is robust to noise compared with the GMM, we attempt to learn image patch priors using the student's-t mixture model (SMM), which is an extension of the GMM. Experiment results demonstrate that our proposed method performs an improvement than the original EPLL.

Keywords: Image denoising · Student's-t mixture model · Expected patch log likelihood · Image patch

1 Introduction

Image restoration plays an important role in image processing and has received much attention recently. Given a corrupted image $u_0 = Au + \varepsilon$, where u is the original image, restoring u from u_0 often is an ill-posed inverse problem and various popular methods have been proposed to solve it.

Generally speaking, there are two classical methods to solve the inverse problem. First, from the perspective of image regularization, estimating u can be solved by means of regularization technique [1]. It has been a well-known and effective tool for solving image inverse problems. Second, from the view of Bayesian statistics, restoring an image means solving the maximum a posterior (MAP) problem [2]. The traditional Bayesian method is mainly composed of the prior probability model and the condition probability model of images. Thus, the key problem of restoring images changes into the learning of image priors. Learning a better image prior is the essence of image restoration. However, due to the high dimensionality of images, learning a good prior from the whole image is a troublesome issue and hard to implement. In order to reduce the computational complexity, patch priors based image denoising method comes into being.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_46

At present, many image restoration methods based on patch priors have been put forward, such as Fields of Experts (FoE) based [3], sparse coding approaches based [4], Expected Patch Log Likelihood (EPLL) based [5], and so on. Among these approaches, the EPLL model is a general optimization framework based on patch priors for image restoration, which restores an image using the Gaussian mixture patch priors learned by Gaussian mixture model (GMM) [6] and is valid and effective for recovering images. However, the GMM is sensitive to noise and can't receive satisfying results when the image is degraded seriously. Recently, the student's-t mixture model (SMM) [7] has been widely used in several fields of image processing because of its robustness to noise. Thus, in this paper, we learn the patch priors by the SMM and then denoise images by maximizing the expected patch log likelihood.

2 Proposed Method

2.1 Background of Expected Patch Log Likelihood

Expected patch log likelihood (EPLL) is a general optimization framework based on patch priors for image restoration, which accumulates patch priors in place of the whole image prior for reducing the computational complexity. Generally, the EPLL is defined as follows:

$$EPLL(u) = \log p(u) = \sum_i \log p(P_i u) \tag{1}$$

where P_i denotes an operator for extracting image patch u_i from image u at position i.

As mentioned above, the MAP is a classical method for image restoration. Given a corrupted image $u_0 = Au + \varepsilon$, restoring u from u_0 by the MAP is equivalent to maximize the following equation:

$$\begin{aligned} \max_u p(u|u_0) &= \max_u \{\log p(u_0|u) + \log p(u)\} \\ &= \min_u \{-\log p(u_0|u) - \log p(u)\} \\ &= \min_u \{-\log p(u_0|u) - \sum_i \log p(P_i u)\} \\ &= \min_u \{-\log p(u_0|u) - EPLL(u)\} \end{aligned} \tag{2}$$

Thus, the optimization framework for image restoration based on EPLL is presented as follows:

$$\min_u \left\{ \frac{\lambda}{2} \|u - u_0\|^2 - \sum_i \log p(P_i u) \right\} \tag{3}$$

where λ is the regularization parameter. This optimization equation can be solved by the alternating minimization (AM) algorithm, which introduces a set of auxiliary variables $\{z_i\}$ and changes the cost function in the following form:

$$\min_{u,\{z_i\}} \left\{ \frac{\lambda}{2} \|u - u_0\|^2 + \sum_i \left\{ \frac{\beta}{2} \left(\|P_i u - z_i\|^2 \right) - \log p(z_i) \right\} \right\} \tag{4}$$

where β is the penalty parameter which often is set to be large enough to ensure that the solution of (4) is close to that of (3). Then Eq. (4) can be minimized by alternatively updating z_i and u.

2.2 Proposed Method Using Student's-t Mixture Patch Priors

The traditional EPLL model restores an image using the Gaussian mixture patch priors learned by the GMM which is sensitive to noise. In this paper, for further enhancing the robustness of the EPLL model for image denoising, we learn the patch priors using the SMM. In general, the density function of the SMM on each patch u_i extracted from an image u is defined as:

$$p(u_i) = \sum_{j=1}^{K} \pi_j S(u_i | \Theta_j) \tag{5}$$

where π_j is the mixing weights, $\Theta_j = \{\mu_j, \Lambda_j, \upsilon_j\}$, μ_j, Λ_j and υ_j are the corresponding mean, precision(inverse covariance matrix) and degree of freedom and K is the number of mixture components.

Because the SMM assumes that image patch is independent of each other, the joint conditional density of the image can be written as:

$$p(u) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j S(u_i | \Theta_j) \tag{6}$$

From Eq. (6), we learn the patch prior by introducing the EPLL and the image restoration model is given by:

$$\min_u \left\{ \frac{\lambda}{2} \|u - u_0\|^2 - \sum_i \log p(P_i u) \right\} \tag{7}$$

The Eq. (7) can be solve using AM algorithm by introducing a set of auxiliary variables $\{z_i\}$, updating z_i and u alternatively as follows:

$$z_i^{n+1} = z_i^n + \Delta t[\lambda(P_i u_0 - P_i u^n) + \frac{(\upsilon_{k_{\max}} + D)(P_i u^n - \mu_{k_{\max}})\Lambda_{k_{\max}}}{\upsilon_{k_{\max}} + (P_i u^n - \mu_{k_{\max}})^T \Lambda_{k_{\max}} (P_i u^n - \mu_{k_{\max}})}] \quad (8)$$

$$u^{n+1} = (\lambda I + \beta \sum_i P_i^T P_i)^{-1} \cdot (\lambda u_0 + \beta \sum_i P_i^T z_i^n) \quad (9)$$

where Δt is the time step, I is the identity matrix.

In summary, our suggested algorithm can be implemented as follows:

Step1. Input corrupted image u_0, model parameters λ and $\beta, \Delta t$ and iteration stopping tolerance ε;
Step2. Calculate z_i^1 using (8);
Step3. Calculate u^1 using (9) with constant regularization parameters λ;
Step4. Calculate z_i^{n+1} using (8);
Step5. Pre-estimate image u^{n+1} using (9)with constant λ;
Step6. Repeat Steps 4−5 until satisfying stopping criterion

3 Experiment Results and Analysis

In our experiments, in order to learn a better patch prior, we train the SMM with 200 mixture components from a set of 2×10^6 images patches sampled from the Berkeley Segmentation Database Benchmark (BSDS300). Comparing our proposed method with the original EPLL, the parameters are set as follows: the image patch size $l = 8$, the weighted coefficients $\lambda = l^2/\sigma^2$, $\beta = 1/\sigma^2 * [1\ 2\ 4\ 8\ 16]$, and the noise standard variance $\sigma = 25$.

Figure 1 demonstrates the denoised results of the EPLL with Gaussian mixture priors and student's-t mixture priors respectively on Test1 image (i.e., No. 40374). The related quantitative comparison, in terms of peak signal to noise ratio (PSNR), is shown in Table 1. From the denoised results shown in Fig. 1(c) and (d), we can see that our proposed method outperforms the original EPLL with Gaussian mixture priors. This is probably because that our method learns image priors using SMM, which has a good robustness to noise. In addition, as demonstrated in Table 1, the PSNR value of our method is higher than the original EPLL. In Fig. 2 we also compare our proposed method with the original EPLL on Test2 image (i.e., No. 160068). From the Fig. 2(d) we can find that our proposed method make the denoised result smoother than the original EPLL.

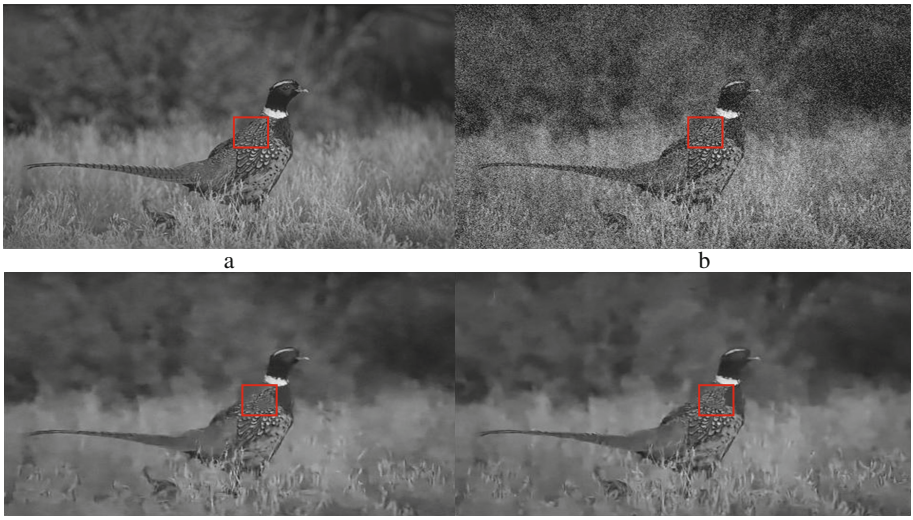


Fig. 1. Denoising results on Test1 image

Table 1. The PSNR results of different denoising models

Image	EPLL	Our method
Test1	29.72	29.93
Test2	28.98	29.12

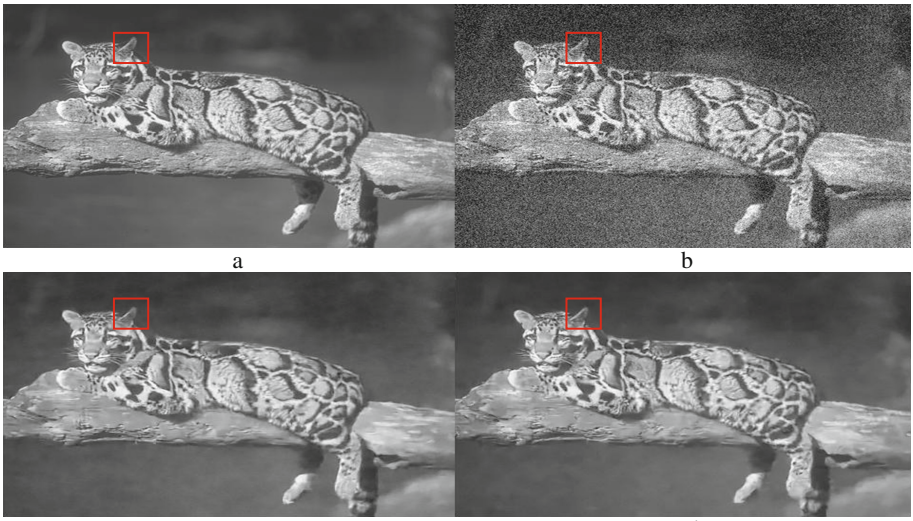


Fig. 2. Denoising results on Test2 image

4 Conclusions

Image denoising method based on patch priors has been a hot issue. Expected patch log likelihood is a general denoising method based on the Gaussian mixture patch priors. In this paper, we use the student's-t mixture model in place of the Gaussian mixture model to learn patch priors owing to the fact that the SMM has a heavy tail and is more robust to noise than the GMM. Our proposed method can learn a better patch prior and obtain more satisfying visual effects in image denoisng, especially when the image is corrupted seriously.

Acknowledgments. This work was supported in part by the NSFC (Grants 61402234 and 61402235) and the PAPD.

References

1. Yang, Z., Jacob, M.: Nonlocal regularization of inverse problems: a unified variational framework. IEEE Trans. Image Process. **22**, 3192–3203 (2013)
2. Chantas, G., Galatsanos, N., Likas, A.: Maximum a posteriori image restoration based on a new directional continuous edge image Prior. In: IEEE International Conference on Image Processing, ICIP 2005, vol. 1, pp. I-941–I-944. IEEE (2005)
3. Roth, S., Black, M.J.: Fields of experts. Int. J. Comput. Vis. **82**, 205–229 (2009)
4. Dong, W., Zhang, L., Shi, G., et al.: Nonlocally centralized sparse representation for image restoration, pp. 1620–1630. IEEE Transactions on, Image Processing (2013)
5. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 479–486. IEEE (2011)
6. Nguyen, T.M., Wu, Q.M.J.: Gaussian-mixture-model-based spatial neighborhood relationships for pixel labeling problem. IEEE Trans. Syst. Man Cybern. Part B Cybern. **42**, 193–202 (2012)
7. Zhang, H., Wu, Q.M., Nguyen, T.M., et al.: Synthetic aperture radar image segmentation by modified student's t-mixture model. IEEE Trans. Geosci. Remote Sens. **52**, 4391–4403 (2014)

Regularization Parameter Selection for Gaussian Mixture Model Based Image Denoising Method

J.W. Zhang[1(✉)], J. Liu[1], Y.H. Zheng[2], and J. Wang[2]

[1] College of Math and Statistic, Nanjing University of Information Science and Technology, Nanjing 210044, China
zhangjw@nuist.edu.cn
[2] CICAEET, College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract. Regularization parameter selection for image denoising has always been a hot issue. In this paper, an adaptive regularization parameter selection method is exploited for the Gaussian Mixture Model (GMM) based image restoration by combining the gradient matching and the local entropy of the image, which varies with different regions of the image and has a good robustness to noise. Experiment results demonstrate that our proposed adaptive regularization parameter for GMM based image restoration method performs comparatively well, both in visual effects and quantitative evaluations.

Keywords: Image denoising · Gaussian mixture model · Image prior · Adaptive regularization parameter

1 Introduction

In recent years, the digital image has been widely used in our daily life. However, during the acquiring process, it is inevitably corrupted by the degraded factors, mainly including the precision in measurements of sensors, motion blur, lens aberration. Therefore, in order to obtain the high quality images, there has been a growing attention in image denoising techniques.

In the past decades, a great variety of image recovery methods have been presented, such as the regularization methods based [1], the image sparse representation based [2], the mixture models learning based [3], and so on. Among them, the mixture models learning based, particularly the Gaussian Mixture Model (GMM) [3] learning based image restoration method has proven its effectiveness and achieved good results.

Recently, regularization parameter selection has received much attention, which has a great effect on preserving more details of image when denoising. At present, numerous methods for regularization parameter selection have been put forward, including the discrepancy principle based [4], the residual image statistics (RIS) based [5], the L-curve and gradient based [6, 7]. In this paper, we focus on the image gradient based regularization parameter selection problem for GMM based image restoration model. Unfortunately, the image gradient is sensitive to noise and can't acquire

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_47

satisfying results when the image is corrupted seriously. By observing that the local entropy of image has a good robustness to noise, we attempt to construct a new adaptive regularization parameter using image gradient matching and the local entropy of image. Moreover, we analyze the relationship between the regularization parameters of the data-fidelity term and gradient-fidelity term, and propose a novel regularization parameter scheme for GMM based image restoration method for preserving more small-scale textures and details of images.

2 Proposed Method

2.1 GMM Based Image Recovery Model

Given an image u with N pixels, let $u_i(i = 1, 2, \ldots, N)$ denote an image patch with the size of $\sqrt{L} \times \sqrt{L}$, obtained by $u_i = R_i u$, where R_i denotes an operator for extracting image patch u_i from image u at position i. The joint conditional density of the image is given by:

$$p(u) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j N(u_i | \mu_j, \Sigma_j) \tag{1}$$

where π_j is the mixing weights, K is the number of mixture components, μ_j and Σ_j are the corresponding mean and covariance matrix.

Then we learn the patch priors using (1) by introducing the EPLL, and the optimization is presented as follows:

$$\begin{aligned} &\min_u \left\{ \tfrac{\lambda}{2} \|u - u_0\|^2 - EPLL(u) \right\} \\ = &\min_u \left\{ \tfrac{\lambda}{2} \|u - u_0\|^2 - \sum_i \log p(R_i u) \right\} \end{aligned} \tag{2}$$

where λ is the regularization parameter.

2.2 Proposed Method with Adaptive Regularization Parameter

Regularization parameter plays an important role in image restoration. As explained in [6, 7], an adaptive regularization parameter for the fidelity term is valid for preserving fine structures of images. In this paper, so as to acquire more satisfying denoised results, a novel Gaussian mixture model based image denoising method with adaptive data-fidelity term and gradient-fidelity term is presented as follows:

$$\min \left\{ \frac{\lambda(x,y)}{2} \|u - u_0\|^2 + \frac{\alpha(x,y)}{2} \|\nabla u - \nabla(G_\sigma * u_0)\|^2 - \sum_i \log p(R_i u) \right\} \tag{3}$$

where ∇ denotes the image gradient, G_σ is a Gaussian filter operator, the second term is the gradient fidelity term, λ and α are the weight coefficients and can be chosen as a function of the gradient and local entropy of corrupted image in the following form:

$$\begin{aligned} \alpha(x,y) &= \frac{1}{1+(|\nabla u|\cdot f(E(x,y)/k_0)^{g(|\nabla u|)}} \\ \lambda(x,y) &= k(1-\alpha(x,y)) \end{aligned} \tag{4}$$

where k is a given constant, k_0 is a threshold value, $g(|\nabla u|)$ is designed by:

$$g(|\nabla u|) = 2 + \frac{k_1}{1+(|\nabla u|/k_0)^2} \tag{5}$$

where k_1 is also a threshold value, when ∇u is large, $g(|\nabla u|) \to 2$, when $\nabla u \to 0$, $g(|\nabla u|) \to 2+k_1$, $E(x,y)$ is the local entropy of image and $f(E(x,y))$ is defined as follows:

$$f(E(x,y)) = 1 + \frac{E(x,y) - \min(E(x,y))}{\max(E(x,y)) - \min(E(x,y))} M \tag{6}$$

where M is the maximum of image gradient norm.

The regularization parameters vary with different regions of the image. $\lambda(x,y)$ is set to be small in the smooth regions of image, while $\alpha(x,y)$ is large so as to remove much noise while guarantee the similarity between the restored image and the corrupted one. At the edge of image, $\alpha(x,y)$ is small but $\lambda(x,y)$ is large, for preserving more edges of the image. Meanwhile, the given k can help balance the regularization parameters $\lambda(x,y)$ and $\alpha(x,y)$, and make them have proper values in different regions of the image for preserving more details of image while smoothing noises.

Here, we employ the Half Quadratic Splitting algorithm to solve (3). The Eq. (3) is equivalently transformed into the following function by introducing a set of auxiliary variables $\{z_i\}$ as follows:

$$\min_{u,\{z_i\}} \left\{ \begin{aligned} &\frac{\lambda(x,y)}{2}\|u-u_0\|^2 + \frac{\alpha(x,y)}{2}\|\nabla u - \nabla(G_\sigma * u_0)\|^2 \\ &+ \sum_i \{\frac{\beta}{2}(R\|Ru-z_i\|^2) - \log p(z_i)\} \end{aligned} \right\} \tag{7}$$

where β is the penalty parameter.

For solving (7), at first, we choose the most likely Gaussian mixing weight $j_{\max}$ for each patch $R_i u$, then Eq. (7) is minimized by alternatively updating z_i and u:

$$z_i^{n+1} = (\Sigma_{j_{\max}} + \frac{1}{\beta} I)^{-1} \cdot (R_i u^n \Sigma_{j_{\max}} + \frac{1}{\beta}\mu_{j_{\max}} I) \tag{8}$$

$$u^{n+1} = u^n + \Delta t[\lambda(x,y)(u_0 - u^n) - \sum_i \beta R_i^T(R_i u^n - z_i^n) + \alpha(x,y)(u_{xx}^n + u_{yy}^n) - \alpha(x,y)(G_\sigma * (u_{0xx} + u_{0yy}))] \tag{9}$$

Where I is the identity matrix, Δt is the time step, $\mu_{j_{\max}}$ and $\Sigma_{j_{\max}}$ are the corresponding mean and covariance matrix with the mixing weight $j_{\max}$.

In summary, our suggested algorithm can be implemented as follows:

Step1. Input corrupted image u_0, model parameters $\beta, \Delta t$ and iteration stopping tolerance ε, initialize regularization parameters λ, α;
Step2. Choose the most likely Gaussian mixing weights $j_{\max}$ for each patch $R_i u$;
Step3. Calculate z_i^1 using (8);
Step4. Calculate u^1 using (9) with updated regularization parameters λ and α;
Step5. Calculate z_i^{n+1} using (8);
Step6. Pre-estimate image u^{n+1} using (9) with updated λ and α;
Step7. Repeat Steps 5-6 until satisfying stopping criterion.

3 Implementation and Experiment Results

In our experiments, the GMM with 200 mixture components is learned from a set of 2×10^6 images patches sampled from the Berkeley Segmentation Database Benchmark (BSDS300). All the images used in our experiments are generated by Gaussian noise with zero mean and standard variance $\sigma = 25$. We compare our proposed method with the original EPLL and the EPLL coupling gradient fidelity term with fixed regularization parameters. The parameters in our proposed method are as follows: the image patch size $\sqrt{L} = 8$, the noise standard variance $\sigma = 25$, the weighted coefficients $\beta = 1/\sigma^2 * [1\ \ 4\ \ 8\ \ 16]$, the size of local entropy of the image is set as 3×3, the constant $k_0 = 1/\sigma^2$, $k_1 = 7$ and $k = L/2\sigma^2$.

Figure 1 demonstrates the denoised results of the original EPLL and our proposed method with adaptive λ, α on Test1 image (i.e., No. 3096). The related quantitative comparison, in terms of peak signal to noise ratio (PSNR) and signal to noise ratio (SNR), are shown in Table 1. The Fig. 1(a) and (b) are respectively the original image and noisy image. The Fig. 1(c) shows that the denoised result obtained by the original EPLL, and we can see that some regions of the image are not smooth. In contrast, the Fig. 1(d) acquired by our method achieves a better result. This is probably due to the fact that our proposed method incorporates the gradient fidelity term with the EPLL, which can help preserve more details of image and make the degraded image smoother during the denoising procedure.

Figure 2 demonstrates the denoised results of our proposed method and the comparison with the EPLL with fixed λ, α on Barbara image. The corresponding PSNR and

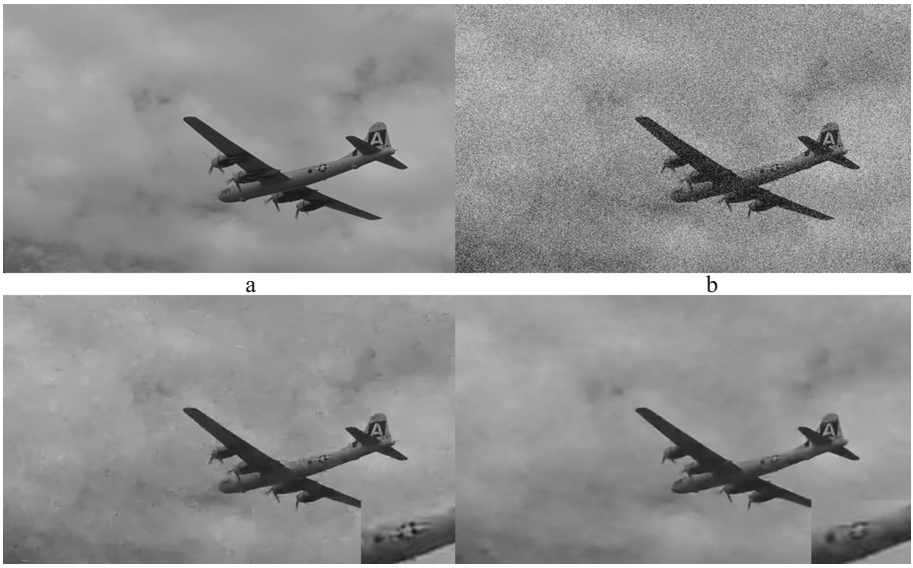


Fig. 1. Denoising results on Test1 image

Table 1. The PSNR and SNR results of different denoising models

Test1	EPLL	EPLL + our adaptive λ, α
PSNR	36.09	36.88
SNR	16.14	16.94

SNR are shown in Table 2. We enlarge the right shoulder of denoised result and put it on the right of image. From the result in Fig. 2(c) we can see that some small-scale textures of the image are not clear, while the result of our proposed method in Fig. 2(d) preserve more textures. This is probably on account of the fact that the parameters λ, α of our proposed method vary with different regions of the image. They can change their values automatically according to the image information and help preserve more fine structures in image. Therefore, by comparing, our proposed adaptive method outperforms the EPLL with a fixed λ, α both in PSNR and SNR.

Table 2. The PSNR and SNR results of different denoising models

Barbara	EPLL + fixed λ, α	EPLL + our adaptive λ, α
PSNR	27.78	27.95
SNR	13.94	14.12

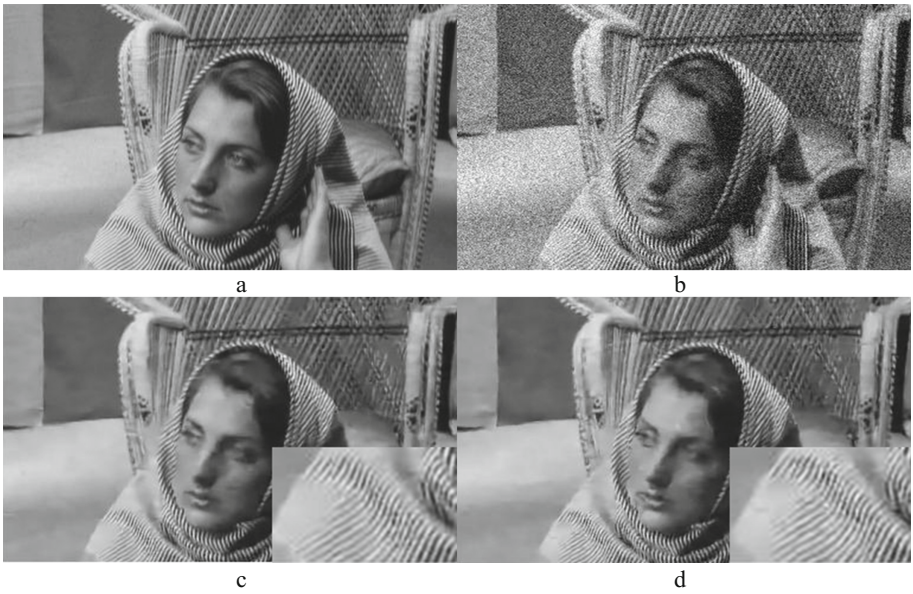


Fig. 2. Denoising results on Barbara image

4 Conclusions

The GMM based image denoising method has received much attention in recent years. In this paper, we devote to the research of the regularization parameter selection for the GMM based image denoising model. We construct an adaptive regularization parameter coupling the local entropy of the image, which varies with different regions of the image and is robust to noise. Our proposed method achieves a satisfying denoised result and shows a clear improvement compared with the original EPLL algorithm in image denoising.

Acknowledgments. This work was supported in part by the NSFC (Grants 61402234 and 61402235) and the PAPD.

References

1. Wen, Y.W., Chan, R.H.: Parameter selection for total-variation-based image restoration using discrepancy principle. IEEE Trans. Image Process. **21**, 1770–1781 (2012)
2. Peleg, T., Eldar, Y.C., Elad, M.: Exploiting statistical dependencies in sparse representations for signal recovery[J]. IEEE Trans. Signal Process. **60**, 2286–2303 (2012)
3. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 479–486. IEEE (2011)

4. Wen, Y.W., Chan, R.H.: Parameter selection for total-variation-based image restoration using discrepancy principle. IEEE Trans. Image Process. **21**, 1770–1781 (2012)
5. Gilboa, G., Sochen, N., Zeevi, Y.Y.: Variational denoising of partly textured images by spatially varying constraints. IEEE Trans. Image Process. **15**, 2281–2289 (2006)
6. Yuan, Q., Zhang, L., Shen, H., et al.: Adaptive multiple-frame image super-resolution based on U-curve. IEEE Trans. Image Process. **19**, 3157–3170 (2010)
7. Xie, C.C., Hu, X.L.: On a spatially varied gradient fidelity term in PDE based image denoising. In: 2010 3rd International Congress on Image and Signal Processing (CISP), vol. 2, pp. 835–838. IEEE (2010)

Restoration Method for Satellite Image Based on Content-Aware Reciprocal Cell Pool

Yuhui Zheng[1(✉)], Xiaozhou Zhou[1], Tong Li[2], and Jin Wang[1]

[1] CICAEET, College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China
zhang_yuhui@nuist.edu.cn

[2] School of Engineering Science, University of Chinese Academy of Sciences, Beijing 100049, China

Abstract. The challenge for the VHR satellite image restoration is how to cope with the simultaneous deblurring and denoising problem. Although the adaptive reciprocal cell (AR-cell) can be used to suppress noise in image restoration, it is an isotropic linear filter. In this letter, we analyze the AR-cell model, and then a piecewise linear version is developed to adapt to different similar image structures, giving rise to an AR-cell bank. Lastly, based on the bank, a group-wise regularization model is introduced for image restoration. Experimental results demonstrate the promising performance of the proposed method.

Keywords: Image restoration · Variational method · Adaptive reciprocal cell

1 Introduction

Traditional satellite image is unavoidably corrupted by blur, noise and aliasing. Thanks to the Time Delay Integration (TDI)-CCD, the modern satellite imaging system can capture image with low level noise, so that its modulation transfer function (MTF) value at Nyquist frequency can be greatly reduced, resulting in almost no aliasing in observed image. Thus, joint deblurring and denoising is one of the urgent VHR satellite image processing issues.

Normally, image restoration is an ill-posed inverse problem [1]. It amounts to estimate an original image X from a measurement:

$$Y = \Delta^{\Gamma}(H(X)) + \eta \tag{1}$$

Where Y stands for a degraded image, Δ^{Γ} is referred to as a sampling operator on a regular sampling grid Γ, H denotes the point spread function whose Fourier transform is related to MTF, and η is referred to as the additive Gaussian noise with zero mean and variance σ^2. For achieving high quality VHR satellite image, it is natural to address the problem with advanced nonlinear regularization techniques, which can regularize unstable solutions. During the past several decades many kinds of image restoration methods have been proposed [2–9], such as image smoothness based [4, 5], nonlocal similarity based [6, 7], sparse representation based [8–10], etc. Although recent nonlocal

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_48

and sparse representation methods [8, 10] can produce relatively good restored results, and some artifacts in resulting images influence their practical applications. No fake image texture is the first requirement of the following satellite image analysis task. By contrast, smoothness prior based methods are still appealing with smooth solutions.

Generally, image smoothness modeling utilizes gradient related priors [4, 5]. Various methods to date were proposed to characterize some gradient statistic distributions. Among them, variational regularization method is one of the most representative methods, which can be formulated into a minimization problem as:

$$\min_{\bar{X}} \varphi(\|\nabla X\|) + \frac{\lambda}{2}\|Y - H(X)\| \tag{2}$$

Where ∇ is the gradient operator, $\lambda > 0$ is a regularization parameter setting the tradeoff between the two terms, and φ is the Euclidean norm of gradient vector ∇X. When $\varphi(s) = s$, Eq. (2) is the famous total variation(TV) model [11]. Though this type of method can preserve discontinuities, it is often sensitive to noise in recovering image. In addition, the traditional method is hard to generate satisfying results, due to the assumption of smooth solution.

However, since the VHR satellite image has low level noise, this fact could be opening a door again for the traditional variational method [4, 5]. Moreover, gradient information matching based models [12–14] recently have obtained better results in image denoising. Intuitively, if the prior information could be introduced into the classical variational restoration method, we would achieve better results. But it is also well known that image gradient information is also highly sensitive to noise. In this paper, we will show that jointing our proposed AR-cell bank and gradient information matching based variational model can overcome the above problems and produce satisfying result for the VHR satellite images.

2 AR-Cell Model

The concept of reciprocal cell comes from the Fourier transform of sampling operator Δ^{Γ}:

$$F\left(\Delta^{\Gamma}\right) = |D^*|\Delta^{\Gamma *} \tag{3}$$

where F denotes the discrete Fourier transform, Γ_* is a dual grid (when an image is sampled on a grid Γ, its Fourier transform is periodic with respect to the grid Γ_*), and D^* denotes the reciprocal cell which is a tile of dual gird Γ_*. In order to measure the effective resolution of satellite imaging system, taking into account noise level, sampling model, and the MTF, Almansa et al. [15] proposed an AR-cell model:

$$D^*(a,b) = \{\xi : a(\xi) < \tau 1 \; and \; b(\xi) < \tau 2\} \tag{4}$$

Where ξ denotes frequency, $a(\xi)$ expresses the relative aliasing, $b(\xi)$ represents the relative noise, $\tau 1$ and $\tau 2$ are the thresholds. The AR-cell model can be regarded as a regularizer to analyze spectrum image, producing an area called AR-cell.

3 Proposed Methods

3.1 AR-Cell Bank

Due to almost no aliasing in the VHR satellite image, here we assume that the VHR satellite imaging system meets the sampling theorem, thus the support of its spectrum image is completely included in D^* and Eq. (4) can be rewritten as:

$$D^*(b) = \{\xi : b(\xi) < \tau\} \tag{5}$$

Equation (5) picks frequency ξ whose corresponding Fourier coefficient is useful. The relative noise [15] was defined as:

$$b^2(\xi) = E\left(|\tilde{\eta}(\xi)|^2\right) \Big/ E\left(\left|M\tilde{X}(\xi)\right|^2\right) \tag{6}$$

Where $\tilde{\eta}$ is referred to as the noise spectrum, $\tilde{X}$ expresses the original spectrum image, M denotes the MTF, and $\mathrm{E}(\cdot)$ is the expectation operator. Computation of the expected value in Eq. (6) depends on accurate noise and original image models. Normally, the noise is supposed to be the Gaussian white noise, but the ideal spectrum image model (hereafter called image model) is unknown.

By taking the empirical image expected values for a large set of natural images, Almansa et al. [15] suggested that the function $|\xi|^{-p}$, could be taken as the image model. Hence, Eq. (4) can be clearly written as:

$$D^*(b) = \{\xi : (\sigma|\xi|^p / M(\xi)) < \tau\} \tag{7}$$

In Eq. (6), the relative noise $b(\xi)$ is proportional to the distance $|\xi|$ from the origin. So it is not difficult to speculate that the AR-cell yielded by Eq. (7) is perfectly round, and could be used as a low pass filter to suppress noise in recovering image. The relative noise [15] was defined as:

$$b^2(\xi) = E\left(|\tilde{\eta}(\xi)|^2\right) \Big/ E\left(\left|M\tilde{X}(\xi)\right|^2\right) \tag{8}$$

The point here is that the above mentioned universal image model implies that the decay rates of Fourier coefficients in different orientations are the same. In other words, the corresponding filter is isotropic, bringing about over-smoothing result. In practice, VHR satellite image contains abundant texture information so that its Fourier coefficients often concentrate on some orientations with high absolute values. In addition, the image also includes large similar spatial structures. But the previous AR-cell model doesn't represent the self-similarity well because of the linear Fourier transform of the whole image. Naturally, piecewise linear AR-cell models could be more rational to adapt to different similar image structures, generating an AR-cell bank.

3.2 AR-Cell Bank Based Restoration Framework

Given the above analysis, we propose a new restoration method. Firstly, an approximation of the image X, denoted by I, is estimated; Then, we group similar image patches in the estimation I together by clustering; Sequentially, the content-aware AR-cell models are obtained by taking the empirical image expected values for each similar group, giving rise to the AR-cell bank; Lastly, based on the bank, we achieve restored image by using a new restoration model.

To estimate the image X, we utilize the traditional AR-cell to extract the effective spectrum image from the corrupted spectrum image $\tilde{Y}$, and then utilize the Wiener filter to deblur the selected spectrum image and obtain the image I. The key observation under the approach is that when noise spectrum is suppressed by the AR-cell to some extent, then we have $\tilde{Y} \approx M \cdot \tilde{X}$. Thus deblurring operation can be directly carried out. Although the estimation I is a restored image, it is an intermediate result. Nonetheless, compared with the universal image model, the estimation is closer to the original image, so that we can employ it to calculate the AR-cell bank.

After that, similar groups are identified by means of image patches matching:

$$dist(i,j) = \left\| P_i(I) - P_j(I) \right\|_2 \tag{9}$$

where $P_i(I)$ denotes extracting the i th image patch of size $\sqrt{n} \times \sqrt{n}$ from the image I in lexicographic order. In our method, once a patch is selected into a certain similar set, it will be deleted from the patch pool for avoiding inconformity. As a result, we have K groups: $\{S_k\}, k \in [1, K]$ and each group records the positions of the similar patches. Then, we take the empirical image expected values for each similar group and further compute each piecewise AR-cell model by using Eqs. (5) and (6).

In restoration, we begin by selecting image patches from the degraded image Y according to the similar sets $\{S_k\}$. For the sake of simplicity, let f and u denote the l th image patch in the k th similar group of image Y and X, respectively, i.e. $f = Y_l^k$, $u = X_l^k$. Then, we can use the variational regularization to reconstruct image patches. As gradient information matching [12–14] recently has drawn great attention in image denoising, inspired by the works, we introduce the gradient-fitting term into the TV model and propose an AR-cell bank based group-wise restoration model as:

$$\underset{u}{Min} \|\nabla u\| + \frac{\lambda}{2} \|\Pr(f - Hu)\| + \frac{\alpha}{2} \|\Pr(\nabla f - \nabla(Hu))\| \tag{10}$$

Where $\alpha > 0$ is the regularization parameter, and Pr denotes projecting image patch into the area of the corresponding AR-cell. The model consists of three terms: the regularization term, data-fitting term, and gradient-fitting term. Since the last two terms in Eq. (10) are both built in the Fourier domain, we utilize the Fast Fourier Transform (FFT) based splitting technology [5] to recover each patch, group by group. Lastly, we aggregate all the reconstructed patches to form a final whole image by averaging. Note that when the third term in Eq. (10) is canceled, we will obtain another restoration model, i.e. AR-cell bank based TV model (ARTV). The model will be evaluated in experiments.

4 Experimental Results and Analysis

In experiments, in order to explore the performance of the added gradient-fitting term, we compare our method with the ARTV method. Moreover, we also compare the two methods with three representative restoration methods, including the TV [5], IDD-BM3D [7], and NCSR restoration method [10]. All the comparisons are performed under Windows7 and MATLAB v8.0. The source codes of the methods were offered by the authors and four test images (Test 1, 2, 3, 4) shown in Fig. 1 are used. The sizes of test images are all 400 × 400. Note that the first two test images are piecewise smooth while the other twos contain more textures. To verify the performance of image restoration, we firstly blur the test images with a known. Then in view of the high SNR characteristics of the VHR satellite imaging system, low level Gaussian noise with standard deviation $\sigma = 5$ is respectively added into the blurred images. The degraded images are shown in Fig. 2. We use the PSNR to evaluate the quality of the restored images.

In our method, all the test images are divided into 9 × 9 patches ($n = 81$) with 8 pixels patch overlapping. During the iteration of our variational model, the settings of the parameters λ and β are set according to the suggestion in [5]. The initial value of β is set to 4, and the terminal value to 2^{20}. λ is set to $0.05/\max(\sigma^2, 10^{-12})$. The value of parameter α is set slightly smaller than that of the parameter λ to avoid producing false textures. Here we set $\alpha = 0.85\lambda$.

Figure 2 shows the restoration results by the five methods. The TV method puts large penalty on high gradient information in order to avoid amplifying noise and obtain acceptable image as possible as it can, which results in over-smoothed result. We can see that the edge structure in the restored image by the ARTV is clearer than that in the result image by the traditional TV method. This is due to that the piecewise linear AR-cell models can adapt to different image contents and capture more beneficial Fourier coefficients so that improve the restored result. Compared with the ARTV,

Fig. 1. The test images. From left to right and top to bottom: images Test1, 2, 3, and 4.

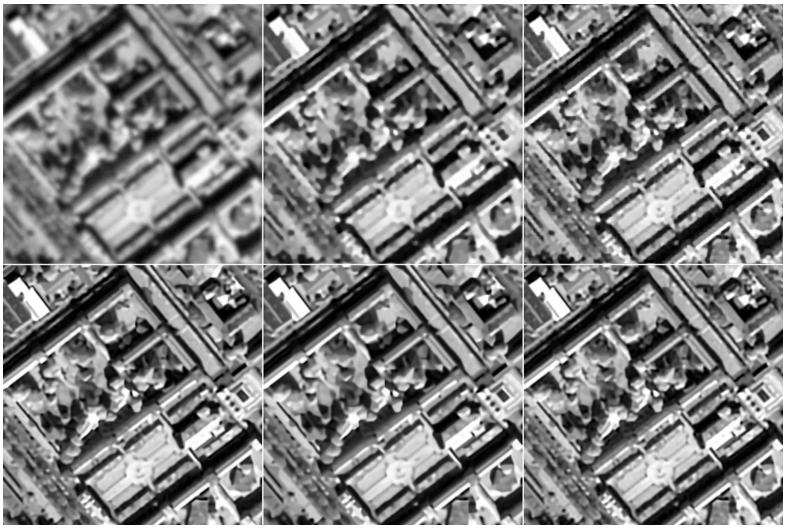


Fig. 2. The performance comparison on the image Test1. From left to right and top to bottom: degraded image ($\sigma = 5$), and restored images by the methods TV, ARTV, IDD-BM3D, NCSR and Our method.

the recovering image by our method is clearer and more natural. One reasonable explanation is that the gradient fitting term has a capability of preserving high frequency information. From an overall perspective, the three methods, i.e. the IDD-BM3D, the NCSR and our method, yield comparatively good images in vision. By observing these images carefully, especially the cars in the streets, we can see that our method can produce slightly better result. Furthermore, Fig. 4 proves the recovering ability of our method. In Fig. 3, we show the average deviations between the original image and the restored images in Fig. 2. Our method gains the smallest average deviation among the five methods.

Table 1 shows that the methods IDD-BM3D and NCSR are highly effective in recovering the texture images, such as the images Test3 and Test4. In Table 1, the average (Avg.) PSNR value of our method is the highest among the five methods. Compared with the state of the art method NCSR, although our method shows a slight advantage in the Avg. PSNR, we can see in Table 2 that our method is on average about 7.5 min (Min.) faster than the NCSR. Our method is more cost-effective. The implement of the IDD-BM3D were by hybrid programming with C language and Matlab, and the other four methods were purely programed with Matlab. In the cause of fairness, we only listed the time consumptions of the other four methods in Table 2. Actually, the IDD-BM3D is also high time consuming [7]. Note that performance comes at a cost of time.

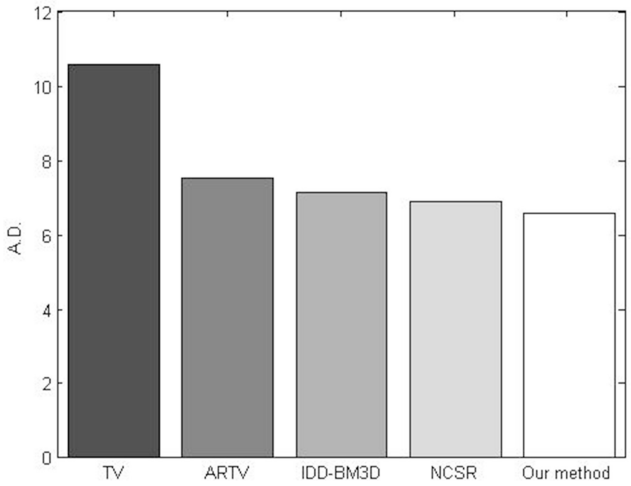


Fig. 3. The comparison of the Average Deviations (A.D.) between the original image and the restored images. TV (A.D. = 10.56), ARTV (A.D. = 7.504), IDD-BM3D (A.D. = 7.107), NCSR (A.D. = 6.980), our method (A.D. = 6.566).

Table 1. Comparison of the SNR (dB) of the five methods

Degraded image	Test1	Test3	Test4	Avg.
	18.12	21.20	21.32	19.67
TV	19.67	22.08	22.11	20.72
ARTV	21.95	23.81	23.94	22.82
IDD-BM3D	22.50	24.22	24.36	23.27
NCSR	22.56	24.26	24.38	23.32
Our method	22.67	24.20	24.35	23.35

Table 2. Comparison of the time consumption (Min.)

Methods	Test1	Test2	Test3	Test4	Avg.
TV	0.033	0.035	0.027	0.025	0.030
ARTV	4.024	4.302	3.615	3.422	3.841
NCSR	12.85	13.42	10.63	9.437	11.58
Our method	4.253	4.528	3.869	3.736	4.097

5 Conclusions

We present a new restoration method by jointing the proposed piecewise linear AR-cell models and a generalization of total variation model coupled with gradient fitting term. The piecewise AR-cell models can adapt to different image structures, yielding the AR-cell bank, which can effectively suppress the noise while extracting useful spectrum information. In addition, the high frequency preservation ability of the gradient

fitting term can further improve the performance of our method. Experimental results prove the effectiveness of our new method.

References

1. Shen, H., Du, L., Zhang, L., Gong, W.: A blind restoration method for remote sensing images. IEEE Geosci. Remote Sens. Lett. **9**(6), 1137–1141 (2012)
2. Krishnan, D., Fergus, R.: Fast image deconvolution using Hyper-Laplacian priors. In: Proceedings of NIPS, pp. 1033–1041 (2009)
3. Zhang, J., Zhong, P., Chen, Y., Li, S.: L1/2-regularized deconvolution network structure of industrial plastics. IEEE Trans. Geosci. Remote Sens. (to be published)
4. Williams, P.: Bayesian regularization and pruning using a Laplace prior. Neural Comput. **7** (1), 117–143 (1995)
5. Wang, Y., Yang, J., Yin, W., Zhang, Y.: A new alternating minimization algorithm for total variation image reconstruction. SIAM J. Imaging Sci. **1**(3), 248–272 (2008)
6. Zhao, M., Zhang, W., Wang, Z., Hou, Q.: Satellite image deconvolution based on nonlocal means. Appl. Opt. **49**(32), 6286–6294 (2010)
7. Danielyan, A., Katkovnik, V., Egiazarian, K.: BM3D frames and variational image deblurring. IEEE Trans. Image Process. **21**(4), 1715–1728 (2012)
8. Mairal, J., Bach, F., Ponce, J., Sapiro, G., Zisserman, A.: Non-local sparse models for image restoration. In: Proceedings of ICCV, pp. 2272–2279 (2009)
9. Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction, Technical report 09-03, Dept. Math., UCLA, Los Angeles (2009)
10. Dong, W., Zhang, L., Shi, G., Li, X.: Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process. **21**(4), 1620–1630 (2013)
11. Bertalmio, M., Caselles, V., Rouge, B., Sole, A.: TV based image restoration with local constraints. J. Sci. Comput. **19**(1–3), 95–122 (2003)
12. Zhu, L., Xia, D.: Staircase effect alleviation by coupling gradient fidelity term. Image Vision Comput. **26**(8), 1163–1170 (2008)
13. Didas, S., Setzer, S., Steidl, G.: Combined l2 data and gradient fitting in conjunction with l1 regularization. Adv. Comput. Math. **30**(1), 79–99 (2009)
14. Zuo, W., Zhang, L., Song, C., Zhang, D.: Texture enhanced image denoising via gradient histogram preservation. In: Proceedings of CVPR, pp. 1203–1210 (2013)
15. Almansa, A., Burand, S., Rouge, B.: Measuring and improving image resolution by adaptation of the reciprocal cell. J. Math. Imaging Vis. **21**, 235–279 (2004)

Student's-t Mixture Model Based Image Denoising Method with Gradient Fidelity Term

J.W. Zhang[1(✉)], J. Liu[1], Y.H. Zheng[2], and J. Wang[2]

[1] College of Math and Statistic, Nanjing University of Information Science and Technology, Nanjing 210044, China
zhangjw@nuist.edu.cn
[2] CICAEET, College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing 210044, China

Abstract. The mixture models based structured sparse representation (MM-SSR) method has received much attention in recent years. Especially, the student's-t mixture model based structured sparse representation (SMM-SSR) has been widely used due to the fact that it has a heavy tail and is robust to noise. In this paper, for further enhancing the performance of SMM-SSR, we attempt to incorporate the gradient fidelity term with the student's-t mixture model for image denoising. Experiment results show that our proposed method outperforms the traditional SMM-SSR method.

Keywords: Image denoising · Student's-t mixture model · Structured sparse representation · Gradient fidelity term

1 Introduction

Image restoration has received much attention in recent years. Generally, restoring an image often is an ill-posed inverse problem, which can be solved by regularization method that is closely related to the image prior modeling. Image priors based image restoration methods have always been a hot issue.

In the past decades, many popular image priors have been presented, such as the gradient based [1], the non-local self-similarity based [2], the sparsity based [3], and so on. Among them, the sparsity prior [3] supposes that image patches can be sparsely represented by a redundant dictionary and it is effective in recovering a wide variety of images. However, the traditional sparse representation method [4] assumes that the atoms in dictionary are independent of each other and the structure information between the atoms is not taken into consideration. In order to achieve more satisfying denoised results, at present, many people have devoted to the research on the correlation among the dictionary atoms and propose a number of structured sparse representation methods, mainly including the nonlocal self-similarity based [2], the group or block-sparsity based [4], and the mixture models based [5, 6]. Thereinto, the mixture models based, especially the student's-t mixture model [7] based structured sparse representation (SMM-SSR) method has been widely used due to the fact that student's-t distribution has heavy tails and a good robustness to noise. Therefore, in this

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_49

paper, we focus on the student's-t mixture model (SMM) based structured sparse representation method for image restoration.

Lately, the gradient information of image has been successfully introduced into a number of models, such as the TV model, the pixel-level MAP-estimation model and non-local sparse representation model and so on. The gradient information of image is valid to help preserve more small-scale textures and details of image. What's more, image gradient has the advantage of acquiring more fine structures of images when denoising. Inspired by the observation, we couple the gradient fidelity term [8] with the SMM-SSR to improve the image restoration performance during the noise removal.

2 Proposed Method

2.1 Expected Patch Log Likelihood Based on SMM

Student's-t mixture model (SMM) has been widely used for learning image patch priors owing to the fact that the Student's-t distribution has a heavy tail and is robust to the outliers. Let $u_i, (i = 1, 2, \ldots, N)$, with dimension D, denote an image patch extracted from an image u, SMM supposes that each patch u_i is independent of each other and the density of an image patch u_i is given by:

$$p(u_i) = \sum_{j=1}^{K} \pi_j S(u_i|\Theta_j) \tag{1}$$

where π_j is the mixing weights, K is the number of mixture components, $\Theta_j = \{\mu_j, \Lambda_j, \upsilon_j\}$, μ_j, Λ_j and υ_j are the corresponding mean, precision (inverse covariance matrix) and degree of freedom, the Student's-t distribution $S(u_i|\Theta_j)$ is defined as:

$$S(u_i|\Theta_j) = \frac{\Gamma(\upsilon_j/2 + D/2)|\Lambda_j|^{1/2}}{\Gamma(\upsilon_j/2)(\upsilon_j\pi)^{D/2}} \times \left(1 + \frac{(u_i - \mu_j)^T \Lambda_j (u_i - \mu_j)}{\upsilon_j}\right)^{-(\upsilon_j + D)/2} \tag{2}$$

where $\Gamma(\cdot)$ is the Gamma function.

Due to the independence of each patch, the joint conditional density of the image can be modeled as:

$$p(u) = \prod_{i=1}^{N} \sum_{j=1}^{K} \pi_j S(u_i|\Theta_j) \tag{3}$$

The Student's-t mixture model can be represented as an infinite Gaussians mixture models and be solved by the Expected Maximum (EM).

Then, the image restoration model based on the expected patch log likelihood (EPLL) is presented as follows:

$$\begin{aligned}&\min_u\left\{\frac{\lambda}{2}\|u-u_0\|^2-EPLL(u)\right\}\\&=\min_u\left\{\frac{\lambda}{2}\|u-u_0\|^2-\sum_i\log p(P_iu)\right\}\end{aligned}\tag{4}$$

where P_i denotes an operator which extracts the i-th patch from image u, λ is the regularization parameter which controls the data fidelity term of the image restoration model.

2.2 Proposed Method with Gradient Fidelity Term

The image gradient has a great effect on preserving fine structures of images when denoising. Motivated by this, we present a novel Student's-t mixture model based image restoration method coupling gradient fidelity term as follows:

$$\min_u\left\{\frac{\lambda}{2}\|u-u_0\|^2+\frac{\alpha}{2}\|\nabla u-\nabla(G_\sigma*u_0)\|^2-\sum_i\log p(P_iu)\right\}\tag{5}$$

where G_σ is a Gaussian filter operator, ∇ denotes the image gradient, λ and α are the parameters that control the weights of each term.

Here, the alternating minimization (AM) algorithm is used to solve the Eq. (5), which converts the optimization into the following form by introducing a set of auxiliary variables $\{z_i\}$:

$$\min_{u,\{z_i\}}\left\{\begin{array}{l}\frac{\lambda}{2}\|u-u_0\|^2+\frac{\alpha}{2}\|\nabla u-\nabla(G_\sigma*u_0)\|^2\\+\sum_i\{\frac{\beta}{2}(\|P_iu-z_i\|^2)-\log p(z_i)\}\end{array}\right\}\tag{6}$$

where β is the penalty parameter with a large value to ensure that the solution of (6) is close to that of (5). For solving (6), we choose the component which has the highest conditional mixing weight $k_{\max}$ for each patch p_iu:

$$\begin{aligned}k_{\max}&=\max_k\ p(k|P_iu)\\&=\max_k\ p(P_iu|k)p(P_iu)\\&=\max_k\ \{\log\ p(P_iu|k)+\log\ p(P_iu)\}\end{aligned}\tag{7}$$

Correspondingly, the image patch P_iu has the mean $\mu_{k_{\max}}$, $\upsilon_{k_{\max}}$ and precision $\Lambda_{k_{\max}}$. Then, Eq. (6) is minimized by alternatively updating z_i and u:

$$z_i^{n+1} = z_i^n + \Delta t[\lambda(P_i u_0 - P_i u^n) + \frac{(\upsilon_{k_{\max}} + D)(P_i u^n - \mu_{k_{\max}})\Lambda_{k_{\max}}}{\upsilon_{k_{\max}} + (P_i u^n - \mu_{k_{\max}})^T \Lambda_{k_{\max}}(P_i u^n - \mu_{k_{\max}})}] \tag{8}$$

$$u^{n+1} = u^n + \Delta t[\lambda(u_0 - u^n) - \sum_i \beta P_i^T (P_i u^n - z_i^n) + \alpha(u_{xx}^n + u_{yy}^n) - \alpha(G_\sigma * (u_{0xx} + u_{0yy}))] \tag{9}$$

where Δt is the time step.

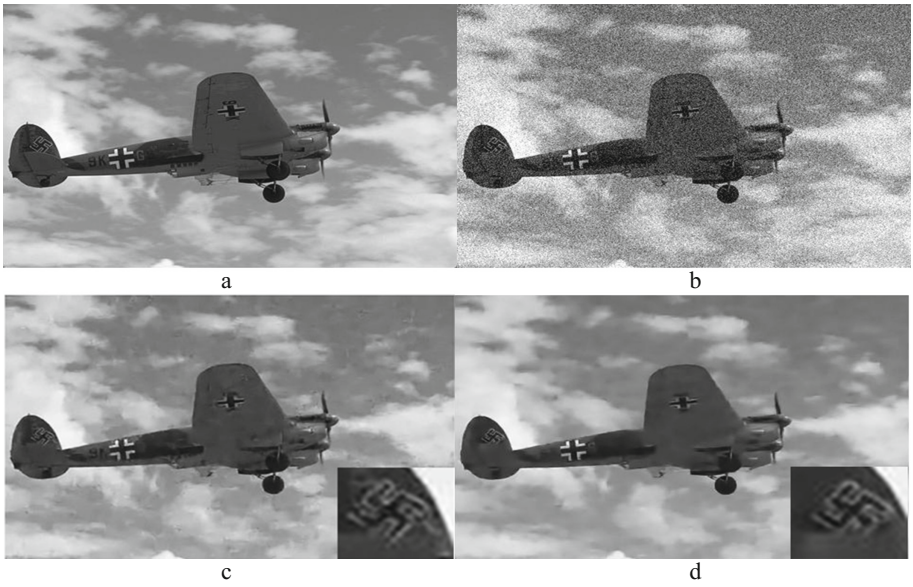
In summary, our suggested algorithm can be implemented as follows:

Step1. Input corrupted image u_0, model parameters λ, α and $\beta, \Delta t$ and iteration stopping tolerance ε;
Step2. Choose the most likely Gaussian mixing weights $k_{\max}$ for each image patch $P_i u$ using (7);
Step3. Calculate z_i^1 using (8);
Step4. Calculate u^1 using (9) with constant regularization parameters λ and α;
Step5. Calculate z_i^{n+1} using (8);
Step6. Pre-estimate image u^{n+1} using (9) with constant λ and α;
Step7. Repeat Steps 5-6 until satisfying stopping criterion.

3 Implementation and Experiment Results

In experiments, we learn the SMM with 200 mixture components from a set of 2×10^6 images patches sampled from the Berkeley Segmentation Database Benchmark (BSDS300). All the images used in our experiments come from the test database of BSDS300 and are generated by Gaussian noise with zero mean and standard variance $\sigma = 25$. To evaluate the performance of our proposed method, we compare our proposed method with the original EPLL, the parameters in our proposed method are set as follows: the image patch size $l = 8$, the weighted coefficients $\beta = 1/\sigma^2 * [1\ \ 4\ \ 8\ \ 16]$, $\lambda = l^2/2\sigma^2$, $\alpha = 1/\sigma^2$, and the noise standard variance $\sigma = 25$.

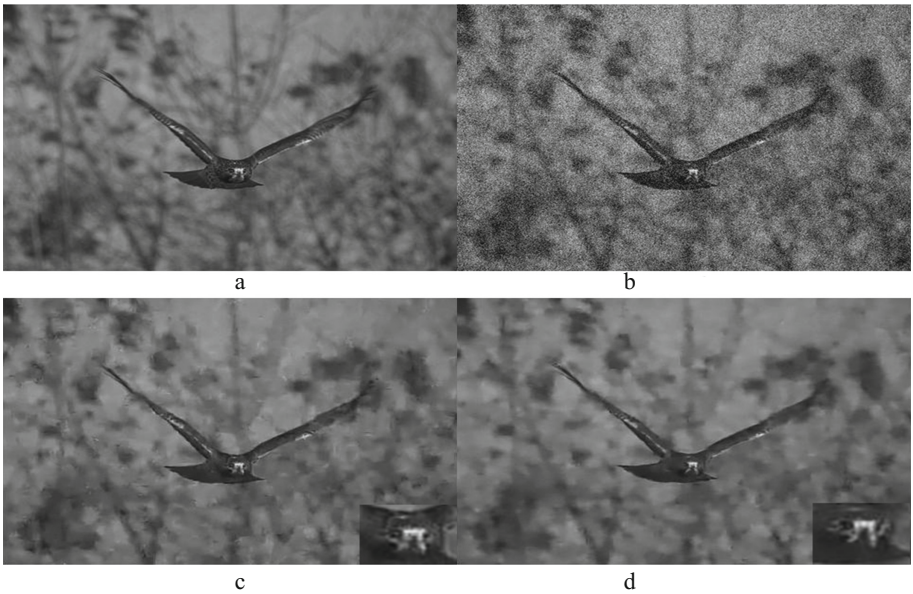
Figure 1 displays the denoised results of the EPLL and our proposed method on Test1 image (i.e., No.3063). The related quantitative comparison, in terms of peak signal to noise ratio (PSNR), is shown in Table 1. From the denoised results shown in Figs. 1(c) and (d), we can see that our proposed method outperforms the original EPLL and make the image smoother. This is because that our method learns image priors using SMM, which has heavy tails and is robust to noise. In addition, we incorporates the gradient fidelity term with the EPLL, which can help preserve more details of image and make the degraded image smoother when denoising. Besides, as demonstrated in



a b

c d

Fig. 1. Denoising results on Test1 image



a b

c d

Fig. 2. Denoising results on Test2 image

Table 1. The PSNR results of different denoising models

Image	EPLL	Our method
Test1	32.84	33.55
Test2	33.21	33.86

Table 1, the PSNR value of our method is higher than the original EPLL. In Fig. 2 we also compare our proposed method with the original EPLL on Test2 image (i.e., No.70011). From the Fig. 2(d) we can find that our proposed method can better preserve the edges and small-scale details of the image. Therefore, by comparison, our proposed method can obtain visually satisfying results and performs better in PSNR.

4 Conclusions

Recently, the mixture models based structured sparse representation (MM-SSR) methods have been popular in image denoising. In this paper, we focus on the research of SMM based image denoising method, which is valid for learning better patch priors. In addition, we couple the image gradient with the SMM-SSR for the sake of preserving more small-scale textures and details during the noise removal. Our proposed method shows a great improvement in peak signal to noise values and achieves more satisfying denoised results.

Acknowledgments. This work was supported in part by the NSFC(61402234, 61402235) and the PAPD.

References

1. Wen, Y., Chan, R.: Parameter selection for total-variation-based image restoration using discrepancy principle. IEEE Trans. Img. Process. **21**, 1770–1781 (2012)
2. Lou, Y., Zhang, X., Osher, S.: Image recovery via nonlocal operators. J. Sci. Comput. **42**, 185–197 (2010)
3. Peleg, T., Eldar, Y.C., Elad, M.: Exploiting statistical dependencies in sparse representations for signal recovery. IEEE Trans. Sig. Process. **60**, 2286–2303 (2012)
4. Zhang, J., Zhao, D., Gao, W.: Group-based sparse representation for image restoration. IEEE Trans. Img. Process. **23**, 3336–3351 (2014)
5. Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 479–486. IEEE (2011)
6. Yu, G., Sapiro, G., Mallat, S.: Solving inverse problems with piecewise linear estimators: From Gaussian mixture models to structured sparsity. IEEE Trans. Img. Process. **21**(5), 2481–2499 (2012)
7. Zhang, H., Wu, Q.M., Nguyen, T.M., et al.: Synthetic aperture radar image segmentation by modified student's t-mixture model. IEEE Trans. Geosci. Remote Sens. **52**, 4391–4403 (2014)
8. Lixin, Z., Deshen, X.: Staircase effect alleviation by coupling gradient fidelity term. Img. Vis. Comput. **26**, 1163–1170 (2008)

Classification Algorithms for Privacy Preserving in Data Mining: A Survey

Sai Ji[1(✉)], Zhen Wang[1], Qi Liu[1], and Xiaodong Liu[2]

[1] College of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China
jisai@nuist.edu.cn
[2] School of Computing, Edinburgh Napier University, 10 Colinton Road, Edinburgh EH10 5DT, UK

Abstract. In the wake of the development in science and technology, consumer or user has produced a large quantity of the information source, whether it is the mobile terminals or the client terminals. When face with such enormous data volume, some people discover the value of the data information for data mining. There are others who initiate be afraid to their privacy information learned or obtained by adversary during data propagation. On top of this, more and more people have undergone threaten of sensitive information loss exactly. Some of them, not only damnify the reputation, but also lose the benefit. Now the challenge is how to balance the security of data and the validity of it. Recently, a large amount of classification algorithms have been applied to process the data to protect data privacy and practicability, such as decision tree, Bayesian networks, support vector machine. In this paper, we overview the learning classification algorithms for privacy preserving in data mining, then make some description with the methods, function, performance evaluation.

Keywords: Data mining · Classification algorithms · Privacy-preserving

1 Introduction

Data mining [1] gets a lot of attention like dig data in recent years. In addition, the application of data mining is more and more widely. It is applied to the field for instance medical treatment, financial sector and stock market. Using data mining may obtain unexpected ponderable results. The user or miner of data may analyze the potential value from the original data source in an even better fashion. Before mining the data, first of all is to preprocess the data which covers data collection, data integration, data reduction, data cleaning and data transformation. The next steps are included in data mining, pattern evaluation and data visualization. The procedure of data mining indicates that the importance of data preprocess, it can be shown as Fig. 1.

Privacy preserving in data mining is a concern to the researcher [2], therefore, improving the technology of the data privacy preserving is the top priority. Moreover, data mining will deal with various types of data like credit card information, identity information and medical history information, which are sensitive information. The sensitive information within the data is likely to bring about someone superfluous

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_50

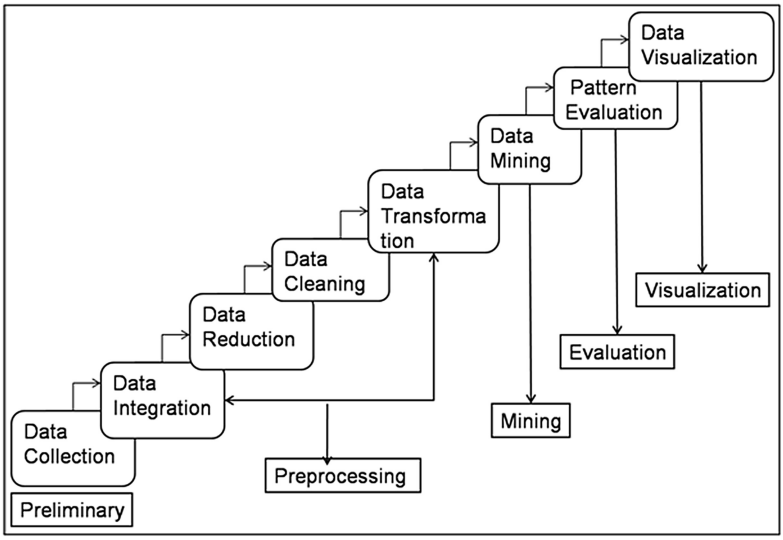


Fig. 1. Data mining steps

endanger. Thus, it should guard the privacy of data in the data mining process. So far, numerous techniques have been applied to the privacy preserving in data mining for classification, clustering, regression [4]. Each one is referred to a number of different classification algorithms in data mining, but classification algorithms play a great role in preserving privacy in data mining.

Classification [1, 3, 4] is a critical technology for privacy-preserving in data mining. Thus, the miners maybe pay more attention to guard the crucial privacy of the owner in the period of classification. Classification is widely used in data classification and predictive information, which implements classifier by learning. Furthermore, the classifier handle the input data through the objective function and then output the classification data, it can be shown as Fig. 2.

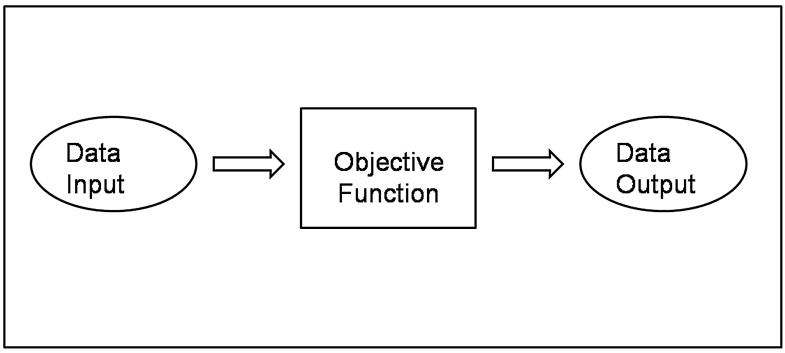


Fig. 2. Classifier model

The paper is made up as follows: In Sect. 2, we review some concepts of support vector machine, making some comparison with different support vector machine as well as analyzing some researches in recent years. In Sect. 3, we demonstrate the definition of Bayesian Classification and protract a flowchart. In Sect. 4, the neural networks will be introduced with the simple notion and model graph. In Sect. 5, we narrate the decision tree for some general knowledge. Finally, a conclusion of this paper is discussed in Sect. 6.

2 Support Vector Machine

2.1 Background Knowledge

Support vector machine is abbreviated as SVM [1], the rest of the writing will use its abbreviation SVM. The SVM is a foremost learning algorithm in machine learning and data mining, which is a supervised learning algorithm [4]. It is popular and high efficiency classification technology applied to many different domains. Furthermore, the SVM is generally used in solving the discrete, high dimensional data and dimensionality suffering during preserving the privacy of information.

The SVM is similar to many other training algorithms with selecting the data to be processed before implement classification [5]. The processed data is divided into two parts, one of them is used in training data set and the other is applied to testing data set. The first step is training and the second step is testing through searching a Maximal Margin Hyperplane (MMH) [6]. The first step: the classifier trains the training data repeatedly that has been marked to acquire the classification parameters. The second step: the classifier tests the testing data set according to the training classification corresponding to the correct class. Besides, the SVM is mainly divided into two segments, the one is liner SVM, the other one is nonlinear SVM, which are used in frequently.

The liner SVM is also called Maximal Margin Classifiers (MMC) [6], then the liner decision boundary and MMC margin will be introduced.

(1) The liner decision boundary can be described with the equation:

$$y = w * x + b \tag{1}$$

w, b of the Eq. (1) are the parameters of model.

(2) The MMC margin is on behalf of the distance of two decision boundary [6], Firstly, assuming two decision boundary are as follows:

$$\begin{aligned} \mathrm{L1} &: y = w * x + b_1 \\ \mathrm{L2} &: y = w * x + b_2 \end{aligned}$$

The MMC margin can be computed with the equation as follows:

$$mar(L1, L2) = \frac{|-b_1 + b_2|}{\sqrt{w^2 + 1}} \tag{2}$$

The nonlinear SVM [5, 6] is a method to use SVM in the nonlinear decision boundary, it can change the original coordinate spaces X into a new coordinate space $\varphi(X)$ and then using the liner decision boundary to realize partition data set easily. The SVM works with the visual graph as shown in Fig. 3.

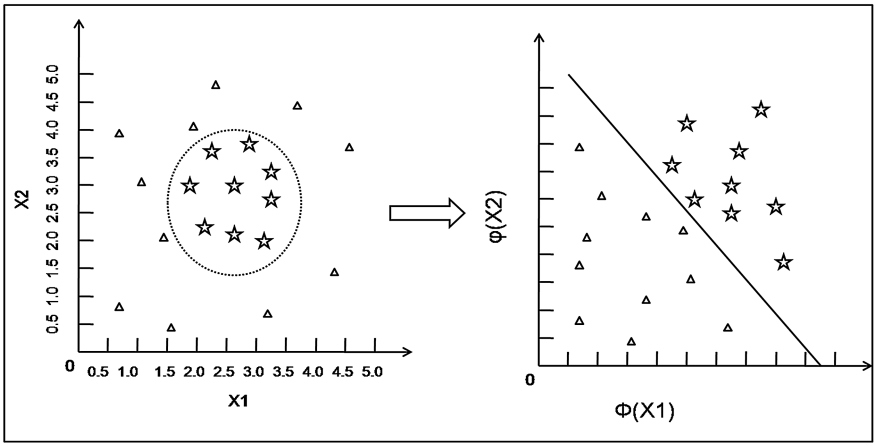


Fig. 3. SVM visualization graph

2.2 Performance Depiction

Many fields are applying the SVM to achieve classification or prediction during preserving privacy. In [5], Rahulamathavan and Phan presented a Client-Server Privacy-Preserving SVM protocol for two-class and multi-class problems through encrypted format for client and the homomorphic encryption properties for server. The protocol is based on a semi-honest models and the information is protected by Paillier homomorphic encryption. In [7], Zhu et al. proposed eDiag framework, the online predictive diagnosis exploits nonlinear SVM with Gaussian kernel, which embeds the random masking technique for user terminal and the polynomial aggregation technique for service provider. In addition, Rahulamathavan and Veluru raised a clinical decision support system by using Gaussian kernel-based classification in [8], the conventional Gaussian kernel-based SVM algorithm is an encrypted-domain algorithm, which is redesigned by using the Paillier homomorphic encryption technique. Sun et al. in [9] formulated a privacy-preserving proximal SVM classification through making use of a global random reduced kernel composed of local reduced kernels with Gaussian perturbations. The employment of formulation yields classification more simple, fast and accurate and improves the performance immensely.

We compare the performance from different point of view for the SVM, which adopted diverse technique. Table 1 makes a comparison on the basic of the method for the state-of-the-art researches. Table 2 distinguishes the function of the techniques attached to disparate area. Table 3 draws a description with various performance evaluations for different papers.

Table 1. The SVM methods

Methods	Types			
	Multi-Class SVM [5]	Nonlinear SVM [7]	SVM [8]	Proximal SVM [9]
Gaussian Kernel	—	√	√	√
Homomorphic encryption	√	√	√	—
Polynomial aggregation	√	√	√	—
Random masking	—	√	—	—

Table 2. The implement function

Rahulamathavan et al. [5]	Data classification, privacy preserving
Zhu et al. [7]	Privacy preserving, predictive diagnosis
Rahulamathavan et al. [8]	Medical forecast, decision support
Sun et al. [9]	More accurate classification

Table 3. Performance evaluation

Types	Performance			
	Accuracy evaluations	Computation time	Computation cost	Communication cost
Multi-Class SVM [5]	√	—	√	√
Nonlinear SVM [7]	√	√	√	√
SVM [8]	√	—	√	√
Proximal SVM [9]	√	√	√	—

3 Bayesian Classification

3.1 Background Knowledge

Bayesian Classification is another algorithm in classification, which is commonly applied to achieve text classification, predictive diagnosis and many other applications. Bayesian classification can greatly deal with the attribute set and class variable relation

modeling [6], which is a method in statistics, a kind of probability as well as statistics knowledge classification algorithm. Bayesian classification can divided into two portions to exposit with training and testing. Moreover, the Bayesian theorem is applied which is a conditional probability about A and B. In the period of training, learning the posterior probability P (B/A) aims at the ensemble learning of A and B. Then the Bayesian tests the testing records. We can explain the posterior probability as the following equation.

$$P(B/A) = \frac{P(A/B)P(B)}{P(A)} \tag{3}$$

The P(B/A) represents the computation of posterior probability, P(A/B) is the meaning of class-conditional probability, P(B) is the prior probability, P(A) is the testimony.

(1) Naive Bayesian Classification is on account of Bayesian theorem and simple probability classifier but highly effective [11]. The process of Naive Bayesian Classification can be learned as shown in Fig. 4.

(2) Bayesian Network expresses a probability relationship between random variable by using figure [4, 6]. Bayesian Network is mainly consisted of directed acyclic graphs (DAG) and homologous probability tables. In [10], the researchers made use of incremental Bayesian Network to preserve the multi-sided sensitive data as well as presented a privacy-preserving protocol. Bayesian Network also is a learning model, which needs to train the data set and then test data set by constructing each time a block of new training data. Updating the existing sufficient statistics of the network structure and then creating a new structure. Next, the probability table of each node could also be computed. The procedure to organize a model of Bayesian Network included as follows, the first step is selecting a related variable set, a variable sequence and assuming the variables are A1, A2,…, An,

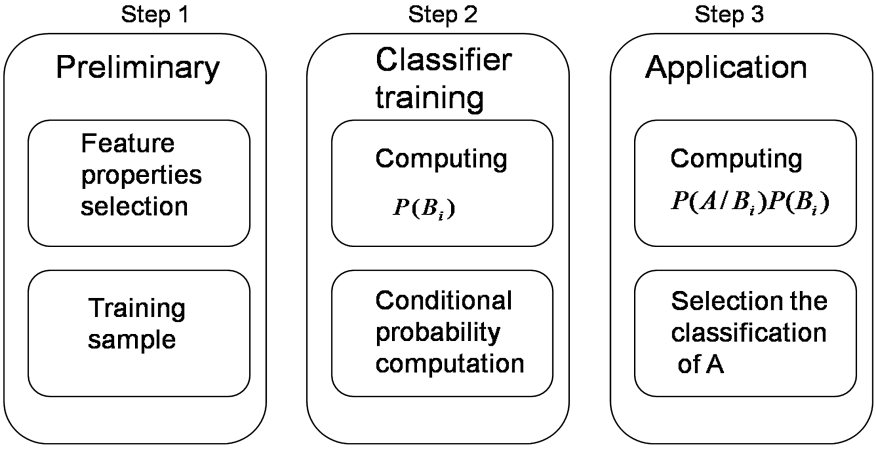


Fig. 4. Naive Bayesian classification

the second step is adding the Ai vertex to the network and set Parent(Ai) to be a minimum subset of A1,…, Ai − 1, such that we have conditional independence of Ai and all other members of A1,…, Ai − 1 given Parents(Ai), the last step is defining the probability table of P(Ai = k/Masks of Parent(Ai)) [10].

A summary will be done about Bayesian Network. The Bayesian Network provides a method to use figure model to capture priori-knowledge and when we build the network successfully is easy to append novel variable which is suitable for to handle half-baked data and has more robust to excessive matching issue.

4 Neural Networks

The thought of neural networks roots in the nerve cells in the organizational structure of human beings is constituted with vast but simple nerve cells and then shaping complex network systems. The neural network also is a highly complicated nonlinear system and applied to classification and prediction for preserving the privacy diffusely in data mining [4]. It is a mathematical model that disposes the distributed parallel information through regulating the conterminous relation of interior massive knots.

4.1 Feed Forward Neural Network

Feed forward neural network is the type of neural network that can be divided into three proportions such as input layer, hidden layer and output layer [4]. Feed forward

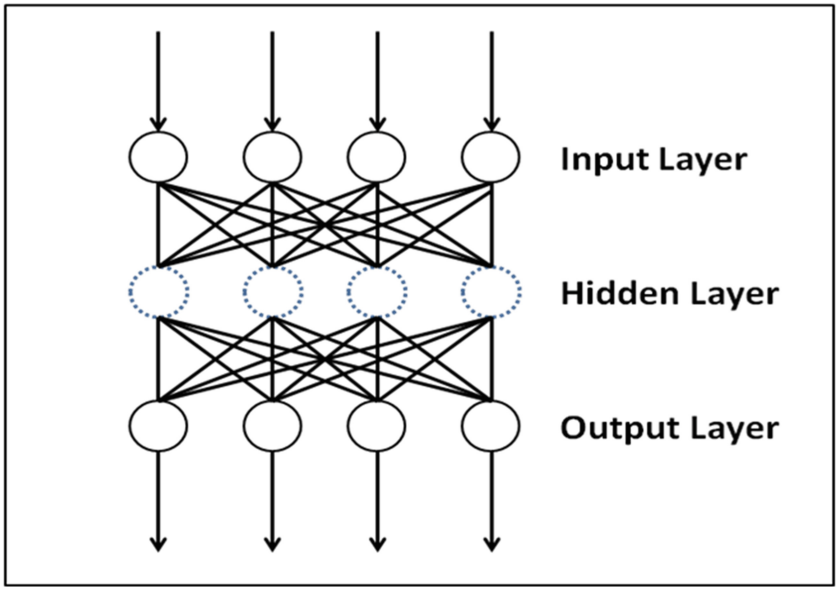


Fig. 5. Feed forward neural network

neural network contains some frequently-used models like perceptron model, back propagation model and radial basis function pattern. Further understanding can be described as Fig. 5.

(1) Perceptron model is ordinary and haploid neural network, mostly applied to pattern classification.
(2) Back propagation (BP) is a multilayer neural network and employs a back propagation to adjust weights. In [12], the authors utilize BP neural network with cloud computing through encrypting the party's private data to bring about preserving the privacy.
(3) Radial basis function (RBF), the hidden layer of RBF is consisted of RBF neuron.

4.2 Feedback Neural Network

Feedback Neural Network is also known as recurrent neural network, the initial status of it is decided by input signal and then reaching equilibrium state under a series of changing condition. The Hopfield neural network is extensive but uncomplicated and easy to coming true associative memory.

4.3 Fuzzy Neural Network

Fuzzy neural network is the vinculum of neural network and fuzzy theory, which can conduct many gradations such as memory, learning, recognition and processing information. There are two manners to comprehend the network. First is a neural network that leads into fuzzy operation and the second is a neural network that is strengthened by the fuzzy logic which achieves the fuzzy system by exploiting neural network architecture. This will enhance interpretability and flexibility of intrinsic network.

5 Decision Tree

5.1 Basic Knowledge

Decision tree is an oversimplified and proverbial tree structure, each internal node indicates the property test, each branch represents the testing output and each leaf node expresses a category [6]. Decision tree is a type of supervised learning that presets a set of samples, each sample has property and category which is determined in advance, and then receives a classifier by learning. The classifier can make the accurate classification for the new input target [4]. Decision tree is structured by training data, which can classify the unknown data efficiently that has some advantages. Then the first one is fine readability and descriptiveness, which contributes to manual analysis, the user need not to comprehend the background knowledge. The second one is high efficiency, which has no use for preparing data or is simple. The decision tree model can build only one time and then employs repeatedly. There are also some disadvantages of decision tree. Firstly, it is more difficult to forecast the continuous data. Secondly, it needs more pretreatment for time sequence data. The decision tree model is shown in Fig. 6.

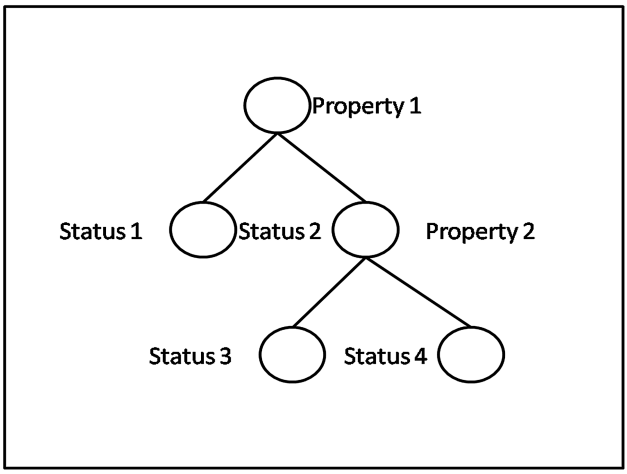


Fig. 6. Decision tree model

5.2 Main Algorithms

(1) Hunt algorithm, it divides the data into subsets by training the records and then establishes decision tree. Assuming that the At the training records set of t and B = {b1, b2,…, bn} is class label. If all records of the At belong to Bt, the t is a leaf node and represents by bt. If not, selecting a attribute testing condition and dividing the records into subsets, founding child nodes for each output and carving up the records of At according to the testing results to child nodes.

(2) Iterative dichotomiser 3 (ID3), it is a greedy algorithm and commonly used in structure a decision tree. It chooses the descent speed of information entropy as the standard to test attributes.

(3) C4.5, it is the extension of ID3 and more efficient. In [13], Behera proposed a method that based on the C4.5 Gini index for conducting distributed data to preserve the multiparty privacy. The different performances of classification algorithms for UCLA Liver Dataset are compared in [14], the results can be known with Table 4 as follows.

Table 4. Performances of classification algorithms

Classification algorithms	Accuracy	Sensitivity
ID3	59.6	56.3
C4.5	68.6	53.1
BP	71.5	57.24
SVM	58.2	68.9

6 Conclusion

Classification is an important technology in data mining. The properties of the classification in data mining include accuracy evaluations, computation time, computation cost and communication cost. The most significant property of the classification is to improve the accuracy and enhance the privacy-preserving in data mining. In this paper, we elaborated the classification algorithms for preserving privacy in data mining. Moreover,we analyzed SVM, Bayesian Classification, neural network and decision tree for different capability. Furthermore, some relevant procedures and comparison diagrams have been described in detail.

Acknowledgement. This work is supported by the NSFC (61300238, 61300237, 61232016, 1405254, 61373133), Marie Curie Fellowship (701697-CAR-MSCA-IFEF-ST), Basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20131004), Scientific Support Program of Jiangsu Province (BE2012473) and the PAPD fund.

References

1. Xu, L., Jiang, C.X., Wang, J., Yuan, J., Ren, Y.: Information security in big data: privacy and data mining. IEEE Access **2**, 1149–1196 (2014)
2. Krawczyk, B., Wozniak, M.: Privacy preserving models of k-NN algorithm. In: Burduk, R., Kurzyński, M., Woźniak, M., Żołnierek, A. (eds.) Computer Recognition Systems 4. AISC, vol. 95, pp. 207–217. Springer, Heidelberg (2011)
3. Lu, W., Jiang, Y.P.: Privacy preserving classification algorithm based on random multidimensional scales. IEEE Comput. Soc. **2**, 406–409 (2009)
4. Kesavaraj, G., Sukumaran, S.: A study on classification techniques in data mining. In: IEEE Communications and Networking Technologies (ICCCNT), pp. 1–7 (2013)
5. Rahulamathavan, Y., Phan, R.C.W., Veluru, S., Cumanan, K., Rajarajan, M.: Privacy-preserving multi-class support vector machine for outsourcing the data classification in cloud. IEEE Trans. Dependable Secure Comput. **11**, 467–479 (2014)
6. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Morgan Kaufmann, San Mateo (2006)
7. Zhu, H., Liu, X., Lu, R., Li, H.: Efficient and privacy-preserving online medical pre-diagnosis framework using nonlinear SVM. IEEE J. Biomed. Health Inf. (2016). doi:10.1109/JBHI.2016.2548248
8. Rahulamathavan, Y., Veluru, S., Phan, R.C.W., Chambers, J.A., Rajarajan, M.: Privacy-preserving clinical decision support system using gaussian kernel-based classification. IEEE J. Biomed. Health Inf. **18**, 56–66 (2014)
9. Sun, L., Mu, W.S., Qi, B., Zhou, Z.J.: A new privacy-preserving proximal support vector machine for classification of vertically partitioned data. Int. J. Mach. Learn. Cybern. **6**, 109–118 (2015). Springer, Heidelberg
10. Samet, S., Miri, A., Granger, E.: Incremental learning of privacy-preserving Bayesian networks. Appl. Soft Comput. **13**, 3657–3667 (2013). Elsevier
11. Huai, M., Wei, L.H., Li, Y., Qi, M.: Privacy-preserving Naive Bayes classification. Knowl. Sci. Eng. Manag. **9403**, 627–638 (2015). doi:10.1007/978-3-319-25159-2_57

12. Yuan, J.W., Yu, S.C.: Privacy preserving back-propagation neural network learning made practical with cloud computing. IEEE Trans. Parallel Distrib. Syst. **25**, 212–221 (2014)
13. Beher, G.: Privacy preserving C4.5 using Gini index. In: Proceedings of 2nd National Conference on Emerging Trends and Applications in Computer Science (NCETACS), pp. 1–4 (2011)
14. Ramana, B.V., Babu, M.S.P., Venkateswarlu, N.B.: A critical study of selected classification algorithms for liver disease diagnosis. Int. J. Database Manag. Syst. (IJDMS) **3**(2), 101–114 (2011)

Exploiting Group Signature to Implement User Authentication in Cloud Computing

Sai Ji[1,2(✉)], Dengzhi Liu[1,2], and Jian Shen[1,2]

[1] Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, China
jisai@nuist.edu.cn
[2] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China

Abstract. Cloud computing is a technology which is developed from the distributed computing. The cloud server provider gathers the redundant storage and computing resource to realize the goal of providing scalable computing resources to consumers. The infrastructures of the cloud computing are virtualized and they can be considered infinite. Therefore, the user side does not need to consider the local storage and computing resource. However, the cloud services are provided by the third party. As a general rule, the user who stores the data in the cloud is not safe. The security of the data is really concerned by the user. In other words, the cloud is interested in the data. We proposed a scheme that can support cloud user's identity authentication, which is based on the group signature. From the security analysis, our scheme can resist some possible attacks.

Keywords: Cloud computing · Security · Identity authentication · Group signature

1 Introduction

Cloud computing is originally developed from distributed computing, it is not a centralized system which has many interconnected computers or servers [1, 2]. Cloud computing offers hardware and software as services which can access data remotely over the internet and it is defined as a large-scale distributed computing paradigm [3, 4].

Due to the concentration of many distributed redundant resources, cloud technology is a thriving tool, which arouses interests of the research fellows and industry domains. A cloud services provider is responsible for providing real time services such as storing the data, giving application and processing the data for consumers. All of the cloud consumers' data or files were remotely outsourced to the cloud servers, these storing data or files can be accessed by the consumers via the internet [5–7]. The consumers can enjoy the cloud services and they will pay for it based on the demand. Hence, ensuring only legal users to access the data is very important in cloud security.

However the cloud is semi-trust because the commercial cloud provider's goal is profit maximization and reduces the computing resource to the users as much as possible [8]. The cloud can collude with the illegal users to access the data [9], thus, the cloud is

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_51

necessary to verify the user and the legal users need to check the accessing servers are the authorized servers, which can avoid the risk of sensitive information leakage.

In recent researches, there are many papers proposed some schemes to solve the security problems in the cloud. In the cloud there are many users' storage data, thus, it is necessary to guarantee that only the data owner and the authorized users can retrieve the data. In order to ensure the privacy of the consumers as well as protect the sensitive data, the data will be outsourced to the cloud in an encrypted form. Moreover, only the privileged users have unrestricted access to user data. Hence, developing a suitable access control scheme is very important. In [10–14] many access control studies have been presented to satisfy the goal of access control in cloud. Using cloud service can decrease the cost of local hardware, software and power consumption. Storing user side's data in the cloud can decrease the local storage burden, especially with the popularity of the portable cloud terminal devices in the future. After data are stored in cloud, or outsourced to the cloud, the data owner don't possess the physical copy of the data. However, the cloud server is semi-trusted and its main goal is to gain benefits, while we can't assure the data we stored is tampered or not. And no matter cloud storage is security or not, most users are always uneasy when they lose their local copy. As a matter of fact, the completeness and correctness of the data that outsourced in cloud is under threat indeed. The threat may result from many aspects, such as the dishonesty of cloud service provider, security vulnerabilities of the cloud server, and so on. Hence, the auditing protocols is needed and in recent cloud security researches there already has many auditing schemes [15–19]. In cloud application a main method to protect the sensitive data is to encrypt data, and then outsource the ciphertext into the cloud sever. When outsource the encrypted data to the cloud, an urgent problem must be faced how to retrieve the encrypted data. Generally the ciphertext cannot provide retrieval semantic and statistical properties, therefore, retrieving the ciphertext is a difficult problem, and the traditional ciphertext retrieval methods cannot fully meet the encrypted cloud data retrieval. The recent main researches on encrypted cloud data search schemes are mainly concentrated on keyword search [20–25].

The direct solution of resisting the data leakage of the cloud is the identity authentication of the consumers. The significant scheme to ensure only the authorized users are able to access the cloud server is that the system can support users authentication. In [26] and in [27] the identity-based authentication schemes are proposed respectively. However, the scheme in [26] is easy to modify the message in transmission process by the adversary and the scheme in [27] is not secure because an attacker may be authenticated successfully with no possession of the password. Therefore, these authentication schemes are not suitable for the cloud environment. In [28] a novel improved user authentication is proposed and the user authentication that can be used in the multi-server circumstances in [29] is proposed. These two authentication schemes can be used in cloud computing. However, the computational cost of them is high.

Due to the amount of the consumers being very large in a cloud, the authorization and authentication are very hard. The main contributions of this scheme can be summarized as follows:

(1) We define a framework to satisfy the cloud authorization and authentication. This framework can support multi-user online identity authentication at the same time.
(2) In order to meet the real cloud environment requirements, our scheme can support dynamic operations, which means the users can join or leave the system at any time.
(3) Due to the features of the technology we adopted, our scheme can resist the impersonating attack. Moreover, our scheme is robust as well as safer than the traditional scheme.

The paper is consisted of following sections. The related works are described in Sect. 2, and some necessary techniques are introduced in Sect. 3. Section 4 introduces the system model and some security definitions. The processes of this scheme are described in Sect. 5, Sect. 6 provided security analysis, followed by a conclusion in Sect. 7.

2 Related Work

2.1 Authentication Scheme

In [26] an identity-based authentication protocol is introduced. It is based on exclusive or and hash operations, thus, it costs less computation resources. It is consisted of registration phase and bidirectional authentication phase.

2.2 Authentication in Cloud

In [29] Yang et al. proposed an authentication protocol which can be used in multi-server cloud circumstances. This scheme consumes less computation costs. This scheme includes registration phase and bidirectional authentication phase.

3 Preliminaries

3.1 Bilinear Maps

Let $\mathbb{G}_1$ and $\mathbb{G}_2$ denote the multiplicative groups of prime order p. The bilinear map $\widehat{e} : \mathbb{G}_1 \times \mathbb{G}_2 \rightarrow \mathbb{G}_T$. We assume $\mathcal{G}_1 \in \mathbb{G}_1$, $\mathcal{G}_2 \in \mathbb{G}_2$ and $\forall a, b \in \mathbb{Z}_p^*$, it may satisfy the properties as follows:

(1) Bilinear: $\widehat{e}(\mathcal{G}_1^a, \mathcal{G}_2^b) = \widehat{e}(\mathcal{G}_1, \mathcal{G}_2)^{ab}$.
(2) Non-degenerate: $\widehat{e}(\mathcal{G}_1, \mathcal{G}_2) \neq 1$.
(3) Computable: $\widehat{e}(\mathcal{G}_1, \mathcal{G}_2)$ can be efficiently computed by an algorithm.

3.2 Complexity Assumptions

The Decision Diffie-Hellman (DDH) Problem: $(\mathbb{G}_1, \mathbb{G}_2, \widehat{e})$ can be described like follows. Giving the random $\mathcal{G} \in \mathbb{G}_1$ and $\mathcal{G}^a$, $\mathcal{G}^b$, $\mathcal{G}^c$ for some $a, b, c \in \mathbb{Z}_p^*$, computing $\widehat{e}(\mathcal{G}_1, \mathcal{G}_2)^{abc} \in \mathbb{G}_2$.

The q-Strong Diffie-Hellman (SDH) Problem: $\mathbb{G}_1$ and $\mathbb{G}_2$ are two multiplicative groups of prime order p, the $\mathcal{G}_1$ and h_1 are generators of $\mathbb{G}_1$ and $\mathbb{G}_2$ respectively. The q-SDH problem is to compute a pair $(\mathcal{G}_1^{1/(\gamma+x)}, x)$ for a given (q + 3)-tuple $(\mathcal{G}_1, \mathcal{G}_1^{\gamma}, \cdots, \mathcal{G}_1^{\gamma^q}, h_1, h_1^{\gamma})$ where γ is randomly generated from $\mathbb{Z}_p^*$. We consider that the q-SDH assumption defines that the adversary A solving the q-SDH problem with negligible advantage.

4 System Model and Security Definitions

4.1 System Model

We designed a cloud system, this system can support common cloud applications and services. The users can join this system to enjoy the system services. The Fig. 1 shows the system model.

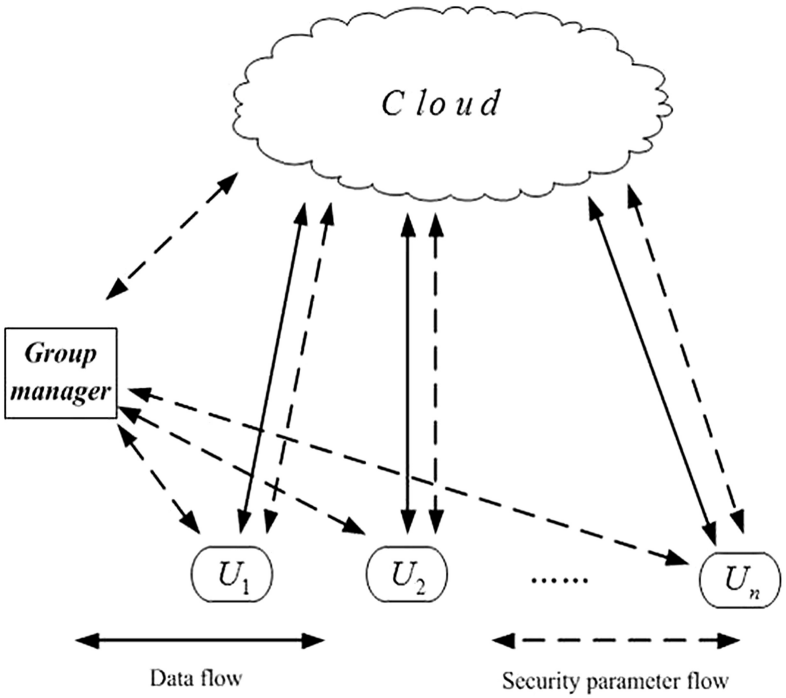


Fig. 1. System model

Cloud server can provide the services to the members, such as storage and computation. It is managed by the cloud server provider, the members cannot take part in the management.

Group manager will help the joining members to register in this system. The group manager as a manager of this system, the services of the cloud is purchased by group manager and it possess the authority to delete and permit the members.

Group members are users in the cloud system, they can enjoy the cloud services, but they can be revoked by the group manager.

4.2 Security Definitions

Cloud server is managed by a cloud server provider, store the data but it is not fully trusted.

Group manager is the administrator of this system. It is fully trusted, but others want attack it to gain some security parameters.

Group members are the users who want to join this system, and they are pass the authentication phase. They are not honest, they can collude with the adversaries.

5 The Proposed Scheme

5.1 Overview

In order to improve the system's security, we design a authentication scheme which is based on the group signature to implement the authentication operation. This scheme includes members' registration and identity authentication. The identity authentication mechanism is transplanted from the group signature scheme [30]. Moreover, there are only one group manager in the group which is responsible for managing the members and ensuring the scheme's security.

5.2 Construction Initialization

Our scheme includes members' identity authentication and group data sharing. The construction is based on the bilinear map. Let $\mathcal{G}$, $\mathcal{G}_1$, $\mathcal{G}_2$ are randomly generated from $\mathbb{G}_1$ and h_1 is randomly generated from $\mathbb{G}_2$. A random parameter $\theta \leftarrow \mathbb{Z}_p^*$ and $h_\theta = h_1$. The group manager needs two private keys $k_1 \in \mathbb{Z}_p^*$, $k_2 \in \mathbb{Z}_p^*$, which are used to re-encrypt the keywords.

5.3 Cloud User Registration

When user is looking forward to joining in the group, it will be regarded as a joining member. Firstly, the joining member sends the JoinRequest to the group manager GM. When the GM receives the JoinRequest, it will randomly generate r_1, a and then compute $A = a^{r_1}$. After this operation it will send the parameter a to the joining

member. We assume the number of the user is k_i, it will generate k_i as its secret key sk and compute a^{k_i} as its public key pk. Then it will send sk and pk to the GM. The GM will compute $P = H(a, pk, A)$ and $s = r_1 + k_i \cdot P(\mathrm{mod}\, p)$. After above operation the member i send (JoinRequest, i) to the GM, the GM receives the request and then calculates $P' = H(a, pk, a^s \cdot pk^{-P})$. The group manager will check whether $P = P'$, if it is true, the group manager will search whether there is $REG[i]$ in the REG list. If there is no $REG[i]$ in the list, the group manager will generate $x, y \leftarrow \mathbb{Z}_p^*$ and then calculate $B = (\mathcal{G}_1 \cdot \mathcal{G}_2 \cdot a^{-sk})^{1/\theta + x}$. At last the group manager will add $REG[i] = (i, \mathcal{G}^y, B, x, y, pk)$ to the REG list. If in the list there exists $REG[i]$, the group manager will locate the $REG[i]$ and extract x,y, then the group manager computes $B = (\mathcal{G}_1 \cdot \mathcal{G}_2 \cdot pk^{-1})^{1/\theta + x}$ like aforementioned operation. At last the group manager will update the $REG[i]$ in the list.

5.4 Cloud User Authentication

There are many users in the cloud data sharing system, we cannot guarantee every user without any malice. In order to improve the security of this scheme, if the user wants to join the group, the group manager will check whether the member has been authorized. If the following Eq. 1 is true, it is mean that the member is an authorized user, otherwise, the group manager will reject the user to join the group, it means this user cannot access the cloud.

$$\widehat{e}\left(B, h_\theta \cdot h_1^x\right) = \widehat{e}\left(\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-sk}, h_1\right) \tag{1}$$

When REG list exist this member, the left formula will be calculated:

$$\begin{aligned} &\widehat{e}(B, h_\theta \cdot h_1^x) \\ &= \widehat{e}((\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot pk^{-1})^{1/\theta + x}, h_\theta \cdot h_1^x) \\ &= \widehat{e}(\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-k_i}, h_1^\theta \cdot h_1^x)^{1/\theta + x} \\ &= \widehat{e}(\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-k_i}, h_1) \end{aligned} \tag{2}$$

When this member cannot find in REG list, the left formula will be calculated:

$$\begin{aligned} &\widehat{e}(B, h_\theta \cdot h_1^x) \\ &= \widehat{e}((\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-sk})^{1/\theta + x}, h_\theta \cdot h_1^x) \\ &= \widehat{e}(\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-k_i}, h_1^\theta \cdot h_1^x)^{1/\theta + x} \\ &= \widehat{e}(\mathcal{G}_1 \cdot \mathcal{G}_2^{-y} \cdot a^{-k_i}, h_1) \end{aligned} \tag{3}$$

6 Security Analysis

Theorem 1: *Our system can support the impersonating attack resistance.*

In this scheme's registration phase, the joining member needs to send a JoinRequest to the group manager, the request must be admitted by the group manager and then this member can be registered successfully. Moreover, the different member's secret key *sk* is different. In scheme's registration phase the group manager randomly generates two parameters r_u and a, which can increase the security of the registration. Through an equation in the authentication phase, the system can guarantee that the member is authorized or not. Therefore, the scheme's identity authentication shows it can resist the impersonating attack.

Theorem 2: *Our system is safer than other systems. (The other systems are performed by hash and XOR operations).*

Our system is based on the bilinear map to implement the authentication, through the complexity assumptions, in mathematics the problem is hard to solve. Hence, our system is difficult to break.

7 Conclusions

We design an authentication scheme that can be used in the cloud user's identity authentication. The security analysis indicates that our system is security. Note that our scheme can support the dynamic group. This proposed scheme is very versatile and can be used in a number of real cloud applications. We believe that our scheme can be used in many cloud applications in real world.

Acknowledgement. This work is supported by the National Science Foundation of China under Grant No. 61300237, No. U1536206, No. U1405254, No. 61232016 and No. 61402234, the National Basic Research Program 973 under Grant No. 2011CB311808, the Natural Science Foundation of Jiangsu province under Grant No. BK2012461, the research fund from Jiangsu Technology & Engineering Center of Meteorological Sensor Network in NUIST under Grant No. KDXG1301, the research fund from Jiangsu Engineering Center of Network Monitoring in NUIST under Grant No. KJR1302, the research fund from Nanjing University of Information Science and Technology under Grant No. S8113003001, the 2013 Nanjing Project of Science and Technology Activities for Returning from Overseas, the 2015 Project of six personnel in Jiangsu Province under Grant No. R2015L06, the CICAEET fund, and the PAPD fund.

References

1. Liu, W.: Research on cloud computing security problem and strategy. In: Proceedings of the International Conference on Consumer Electronics, Communications and Networks (CECNet), pp. 1216–1219 (2012)

2. Sabahi, F.: Cloud computing security threats and responses. In: Proceedings of the International Conference on Communication Software and Network (ICCSN), pp. 245–249 (2011)
3. Yang, J., Chen, Z.: Cloud computing research and security issues. In: Proceedings of the International Conference on Computational Intelligence and Software Engineering (CISE), pp. 1–3 (2010)
4. Foster, I., Zhao, Y., Raicu, I., Lu, S.: Cloud computing and grid computing 360-degree compared. In: Grid Computing Environments Workshop, pp. 1–10 (2008)
5. Yang, C., Zhang, W., Xu, J., Xu, J., Yu, N.: A fast privacy-preserving multi-keyword search scheme on cloud data. In: Proceedings of the International Conference on Cloud Computing and Service Computing, pp. 104–110 (2012)
6. Wang, B., Li, B., Li, H.: Oruta: privacy-preserving public auditing for shared data in the cloud. In: Proceedings of the International Conference on Cloud Computing, pp. 295–302 (2012)
7. Zittrower, S., Zou, C.C.: Encrypted phrase searching in the cloud. In: Communications and Information System Security Symposium, pp. 764–770 (2012)
8. Ateniese, G., Burns, R., Curtmola, R., Herring, G., Kissner, L., Peterson, Z., Song, D.: Provable data possession at untrusted stores. In: ACM Conference on Computer and Communications Security, pp. 598–609 (2007)
9. Wang, C., Ren, K., Lou, W., Li, J.: Toward publicly auditable secure cloud data storage services. IEEE Netw. **24**, 19–24 (2010)
10. Ruj, S., Stojmenovic, M., Nayak, A.: Decentralized access control with anonymous authentication of data stored in clouds. IEEE Trans. Parallel Distrib. Syst. **25**, 384–394 (2014)
11. He, H., Li, R., Dong, X., Zhang, Z.: Secure, efficient and fine-grained data access control mechanism for P2P storage cloud. IEEE Trans. Cloud Comput. **2**, 471–484 (2014)
12. Tang, Y., Lee, P.P.C., Lui, J.C.S., Perlman, R.: Secure overlay cloud storage with access control and assured deletion. IEEE Trans. Dependable Secure Comput. **9**, 903–916 (2012)
13. Yan, Z., Li, X., Wang, M., Vasilakos, A.: Flexible data access control based on trust and reputation in cloud computing. IEEE Trans. Cloud Comput. (2015). doi:10.1109/TCC.2015.2469662
14. Wan, Z., Liu, J., Deng, R.H.: HASBE: a hierarchical attribute-based solution for flexible and scalable access control in cloud computing. IEEE Trans. Inf. Forensics Secur. **7**, 743–754 (2012)
15. Yang, K., Jia, X.: An efficient and secure dynamic auditing protocol for data storage in cloud computing. IEEE Trans. Parallel Distrib. Syst. **24**, 1717–1726 (2013)
16. Yuan, J., Yu, S.: Public integrity auditing for dynamic data sharing with multiuser modification. IEEE Trans. Forensics Secur. **10**, 1717–1726 (2015)
17. Zhu, Y., Ahn, G., Hu, H., Yau, S.S., An, H.G., Chen, S.: Dynamic audit services for integrity verification of outsourced storages in clouds. In: ACM Symposium, pp. 1550–1557 (2011)
18. Li, J., Tan, X., Chen, X., Wong, D.S., Xhafa, F.: OPoR: enabling proof of retrievability in cloud computing with resource-constrained devices. IEEE Trans. Cloud Comput. **3**, 195–205 (2015)
19. Wang, B., Li, H., Liu, X., Li, F., Li, X.: Efficient public verification on the integrity of multi-owner data in the cloud. J. Commun. Netw. **16**, 592–599 (2014)
20. Li, H., Liu, D., Dai, Y., Luan, T.: Enabling efficient multi-keyword ranked search over encrypted mobile cloud data through blind storage. IEEE Trans. Emerg. Top. Comput. **3**(1), 127–138 (2015)

21. Zhang, W., Lin, Y., Xiao, S., Wu, J., Zhou, S.: Privacy preserving ranked multi-keyword search for multiple data owners in cloud computing. IEEE Trans. Comput.(2015). doi:10.1109/TC.2015.2448099
22. Fu, Z., Sun, X., Liu, Q., Zhou, L., Shu, J.: Achieving efficient cloud search services: multi-keyword ranked search over encrypted cloud data supporting parallel computing. IEICE Trans. Commun. **98**, 190–200 (2015)
23. Xia, Z., Wang, X., Sun, X. Wang, Q.: A secure and dynamic multi-keyword ranked search scheme over encrypted cloud data. IEEE Trans. Parallel Distrib. Syst. (2015). doi:10.1109/TPDS.2015.2401003
24. Yu, J., Lu, P., Zhu, Y., Xue, G., Li, M.: Toward secure multikeyword top-k retrieval over encrypted cloud data. IEEE Trans. Dependable Secure Comput. **10**, 239–250 (2013)
25. Li, K., Zhang, W., Yang, C., Yu, N.: Security analysis on one-to-many order preserving encryption-based cloud data search. IEEE Trans. Inf. Forensics Secur. **10**, 1918–1926 (2015)
26. Wang, Y., Liu, J., Xiao, F., Dan, J.: A more efficient and secure dynamic id-based remote user authentication scheme. Comput. Commun. **32**, 583–585 (2009)
27. Das, M.L., Saxenam, A., Gulati, V.P.: A dynamic ID-based remote user authentication scheme. IEEE Trans. Consum. Electron. **50**, 629–631 (2004)
28. Yang, J., Lin, P.: An ID-based user authentication scheme for cloud computing. In: Proceedings of the Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 98–101 (2014)
29. Yang, J., Chang, Y., Huang, C.: A User authentication scheme on multi-server environments for cloud computing. In: Proceedings of the International Conference on Information, Communications and Signal Processing (ICICS), pp. 1–4 (2013)
30. Hwang, J.Y., Chen, L., Cho, H.S., Nyang, D.H.: Short dynamic group signature scheme supporting controllable linkability. IEEE Trans. Inf. Forensics Secur. **10**, 1109–1124 (2015)

A Secure System Framework for an Agricultural IoT Application

Hao Wu, Fangpeng Chen, Hanfeng Hu, Qi Liu, and Sai Ji(✉)

Nanjing University of Information Science and Technology,
No. 219, Ningliu Road, Nanjing 210044, Jiangsu, China
Jisai@nuist.edu.cn

Abstract. The applications of Internet of Things (IoT) permeate to every walk of life due to the development of cheap and efficient electronic components and the increasingly ubiquitous network. Thus it is possible to apply the IoT technologies into agriculture commencing an automated farming method. This paper proposes an IoT framework that can be used to assist in traditional agricultural planting. It integrates two major functions including environment data sensing by a wide variety of sensors and environment factors control with some mechanics driven by intelligent circuits. For example, soil moisture can be monitored and controlled by turning on/off sprinkling irrigation equipment when monitored factors are out of a certain range. In addition, this framework is realized at low cost and easy to deploy. A light-weight data encryption transmission solution based on SSL at the TCP/IP layer is designed to ensure the security of data transmission. A practical testbed based on Ali-Cloud is established to carry huge load of communication and data from an actual deployed agricultural IoT application.

Keywords: Internet of Things · Agricultural automation · Communication security · Sensor network

1 Introduction

Sensing environment data and automatic controlling related devices like irrigation equipment, greenhouse shutter according to a particular program automatically is an efficient way to help farmers to monitor and manage their plantings. The IoT framework of agricultural automation which this paper proposed gives an effective method to make this vision become a reality.

In general, the framework is composed with two parts. One is the part of hardware and another is interface. The hardware's major tasks are sensing and collecting the environment data or other related data to upload the remote server and controlling the related control mechanical equipment according the command download from the server. In order to save costs, the hardware is only responsible for the data collecting and data transmission. And they should be designed to be simple, cheap and well protected for they are usually deployed in a relatively harsh environment like fields where the humidity is too high to the electronic equipment. The interface has much more tasks like receiving the data from hardware which includes decrypting and

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_52

decompressing the data, analyzing and determining whether the data is in accordance with the requirements of the preset and feeding back the commands to the hardware in the fields to take some actions to affect the environment in order to meet the preset value. We think the communication security between the hardware and the remote server is a very important and necessary issue to be considered. At the same time, the huge amount of communication traffic has become concern. As far as possible to reduce the amount of data each communication and saving the communication costs, the data which we transmit should be short as far as possible. So the data that the framework transmits is compressed before encrypting.

In addition, we have applied the framework into an actual project which we call it greenhouse environment data collecting and controlling system (GEDCCS). This project is designed for a greenhouse to monitor its air temperature, air humidity and soil moisture for all day. When the observed values from the sensors are out of the preset threshold ranges, the related controlling devices will do some actions. For example, when the air temperature sensors sense the air temperature lower than the preset air temperature threshold range, the system will control the greenhouse shutter and make it open to let the sun shine in order to increase the greenhouse's temperature to reach the preset value. Besides this, users, the farmers, can set the relevant threshold range according to which kind of crop plants planted in the greenhouse from web dashboard or mobile apps. And then, the greenhouse will keep its air temperature, air humidity and soil moisture in a relatively stable range automatically which greatly reduced the need for human intervention when the amount of greenhouses is too large. At the same time, the users can be clear at a glance of the greenhouses' different environment data which can help the users to make decisions.

2 Challenges and Issues

Because the sensor node equipment of the proposed framework will be deployed in the fields in a large amount, it should be designed to be low cost and easy to use. And the server needs to be able to receive the data from so many sensor nodes at the same time. How to storage so much data is also a hard question to be solved. Last but not least, the security issues will always be an inevitable problem in the Internet of Things.

Low Coat. It is unworthy that to deploy too expensive devices in the fields in a large scale. The agricultural planting's value density is relatively low and the planting area is usually huge. So the hardware which deployed in the fields should be as cheap as possible and has little consumption of the fields' profit. In the greenhouse project mentioned above, each sensor node we developed only costs less than 30$ in hardware. It is just a combination of several modules, which are air temperature sensor, air humidity sensor, soil moisture sensor and SIM card communication module.

Easy to Use. Users are usually farmers who are not professional in electronic engineering. It is important to make the sensor node to be deployed easily and quickly without too much maintenance in later use.

High Performance for High Concurrency and Large Data Volume. It is obviously that the amount of the communication traffic is huge. We can assume that data for each communication is 1 KB, and the frequency of communication is once per second, so the total communication amount is about 84 MB everyday per equipment. And if 1000 equipment has been deployed, the amount of data of each day will be about 82 GB. In fact, the 1000 equipment is far from meeting the needs of market. When the quantity of equipment is too much, the concurrency of communication will be another thorny challenge. Therefore, it is quite a big data issue to be solved both in data receiving and data storage.

Security. The communication of this framework is based on TCP/IP where each message is sent as TCP packets. [1, 2] It is easy for the hackers to intercept the messages while the messages transmitted in network. What's more, they may do some replay attacks or falsify messages if they decrypt the messages and then they will control the equipment in the fields arbitrarily or even maliciously. Therefore, it is necessary to take some encryption actions to prevent network attacks, or reduce the risk of being attacked at least. However, we have required that the hardware design should be simple and cheap, which means that its computing power and storage capacity will be very limit. How to design a light-weight encryption communication solution with these limited resources will be another difficult problem.

3 Blueprint and Framework

The Fig. 1 shows the basic structure of GEDCCS. The left five icons represent the sensors which deployed in the fields. They can sense the different environment data and then send them through the 3G/4G mobile cellular network with the SIM module. The data is transmitted through the Internet and arrives at the server client. In order to be able to withstand high concurrent data transfer requests, they will be diverted into the server pool. The servers in the server pool do the same tasks so that there is no difference in the results of any one server processing. After the server pool's processing, the data will be saved in the database cluster. The web server offers the web service to users so that they can change some settings of their sensor nodes at anytime

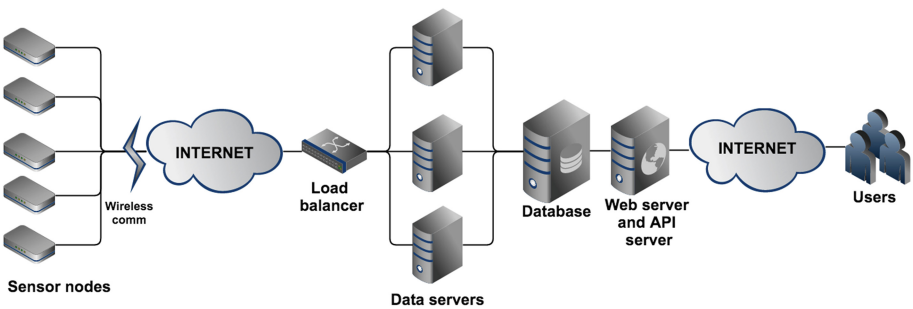


Fig. 1. The basic structure of GEDCCS

and anywhere. Meanwhile, the web server also offers a set of RESTful APIs for mobile apps like android or iOS devices to invoke.

The control commands' distribution is a reverse progress of data uploading. In one hand, the server will determine whether the uploaded sensor data meets the preset threshold which saved in the database. If the value is out of range, the server will construct control commands and distribute them through the Internet to the sensors in the fields. On the other hand, the users can change the related value or settings saved in the database through the web dashboard or the mobile apps. They will take effect when the server queries the database next time.

4 The Hardware Design

The hardware of sensor node is combined with an air temperature and air humidity integrated sensor (SHT10), a soil moisture sensor (self-made) and a SIM GPRS (SIM900) communication module. Besides this, it also has a control module which can control the mechanical devices through the electric relays. All the information included will be processed in the SoC (STM32F102). The Fig. 2 shows the major electronic components which the sensor node used.

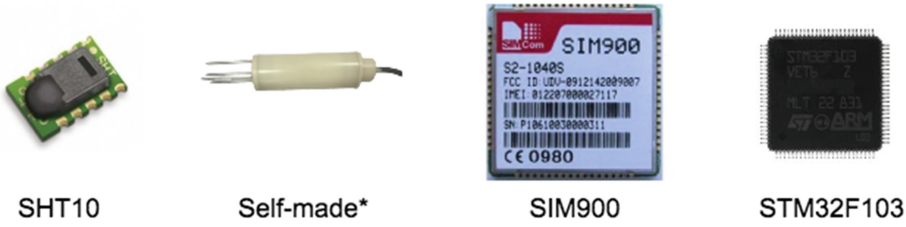


Fig. 2. The major electronic components

This sensor node hardware system can collect the data of air temperature, air humidity and soil moisture in real time, send them to the remote server and automatically carry out irrigation and spray according to the preset threshold.

The SHT100 communicates with the STM32 through the digital serial port. The soil moisture sensor uses analog interface to realize the communication with the STM32. The collecting time and collecting frequency should be set in the STM32. After the STM master chip having collected the data of the sensors and finished the wave filtering and other specific handling, the data will be sent to the remote server which specified with an IP address through the wireless channel with the SIM900 communication module. The SIM900 module and STM32 adopt the way of serial data transmission. The Fig. 3 shows the general construction of the sensor node equipment.

The thresholds of the control like the air temperature, air humidity and soil moisture are set inside the STM32. These thresholds can be updated in real time by the remote terminal such as web dashboard or mobile apps in the help of the SIM900 communication module. Automatic irrigation and sprinkler system is realized by controlling

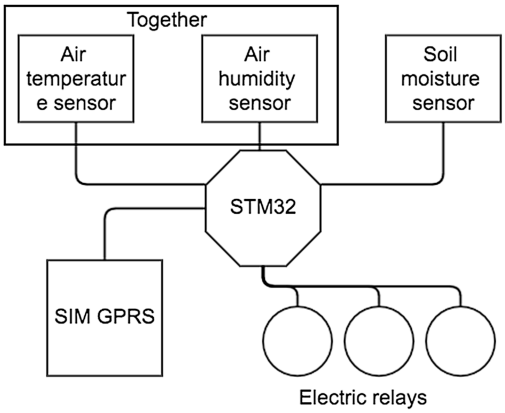


Fig. 3. The general construction of the sensor node equipment

the power supply of irrigation and spray system. The output ports of STM32 can drive the direct-current electric relays by outputting the different high and low levels of pulses, and then realize the function of automatic control.

The price of each SHT10 is about $5, the self-made soil moisture sensor is about $5, the SIM900 module is about $10, the STM32 is about $3, the direct-current electric relay is about $2. Thus, the total hardware cost is no more than $30.

5 The Interface Design

The interface design has two aspects. One the one hand, it is the interaction between the sensor nodes and the remote server which includes the receiving the data from the sensor nodes and sending data to them at the specific time. On the other hand, it is the interaction between the users and the server or sensor nodes which includes showing users the system information and giving users to change the system parameters and control the sensor nodes deployed in the fields. These interactions can be summarized as machine to machine and the machine to human.

Elastic and Light-Weight Communication. The types of sensors are diverse and the communication resources like the bandwidth, the traffic cost, are limited so that they need to save as much as possible. How to define en efficient message format to meet the needs of various types of sensors and make the length of message be as short as possible at the same time is not a simple problem. In this paper, we proposed an elastic and light-weight message format to solve the problem.

At first, we transmit the data in the form of JSON like the following lines:

```
{
  "id": "1001",
  "air_temperature":"20",
  "air_humidity": "70",
  "soil_moisture": "50",
  "time": "20151101083452"
}
```

This message will be transmitted for about once per second, and the useful information are “1001”, “20”, “70”, “50”, “20151101083452”, while the other chars are wasted for they are always transmitted every time. The useful information (24 chars) only accounts for 22 % of the total amount of information (109 chars). It’s obviously that the information density is very low. Even if the JSON strings have a good readability, it’s not suitable to apply in the high frequency data transmission scenario with the limits of bandwidth and traffic cost.

In order to solve the problem, we divide the data transmission process into two steps:

- Step 1: The data sender (sensor node) will send a message format annotation to the data receiver (server), which is just like the following lines:

```
{
  "type":"sensor_a",
  "annotation":
    {
       "id":"0.3",
       "air_temperature":"4.5",
       "air_humidity":"6.7,
       "soil_moisture":"8.9",
       "time":"10.23"
    }
}
```

This JSON string is the SensorA’s format annotation. The type attribute represents that this annotation belongs to SensorA. The annotation attribute’s value is another JSON string which means that the id field will occupy 4 character width from the position 0 to position 3 (Start the index from 0), the air temperature field will occupy 2 character width from the position 4 to position 5, the air humidity field will occupy 2 character width from the position 6 to position 7, the soil moisture field will occupy 2 character width from the position 8 to position 9 and the time field will occupy 14 character width from the position 10 to position 23.

- Step 2: The data sender will only send the data type and data content like the following JSON string:

```
{" sensor_a ":"100120705020151101083452"}
```

The server will choose the SensorA's format annotation to decomposition the message content (100120705020151101083452) with definition of the format sensor_a.

With these two steps, we succeed in extracting redundant information and only send them for once. The rest useful information will be efficiently transferred to the server with a little cost of bandwidth and traffic cost. Comparing with the original long JSON string, the same amount of information will consume only about 20 % of the volume of traffic. This is light-weight. On the other hand, if a new kind of equipment, for example SensorB, wants to access the system, it only needs to send its special data format annotation to the server at first. Then the server will know how to understand the following data sent from this kind equipment. In this way, the system can be compatible with all kinds of equipment only if the equipment realizes this data transmission process. This is elastic.

Data Security and Protection. In order to achieve the goal of safe and reliable communication, we designed a communication process as follows. This process is a reference to the SSL communication mode. It can be seen as a simplified version of SSL. The general idea of this method is to distribute the symmetric key via asymmetric key. The asymmetric encryption algorithms, such as RSA, ECC, have a very high encryption intensity. However, the complexity of the asymmetric algorithms is so high that they are not suitable for devices with limited computing resources. Therefore, we just use them in the process of symmetric key distribution where just a little but very important data needs to be encrypted.

The Fig. 4 shows the main encryption communication flow:

- Step 1: The sensor node initiates the request through establishing a TCP connection at first.
- Step 2: The server will generate a pair of public key and private key stored in the server after receiving the request. Then, the public key will be sent to the sensor node through the TCP connection just established.
- Step 3: The sensor receives the public key. Then it will generate a symmetric key randomly for the next series of communications. The symmetric key will be sent to the server after being encrypted with the public key received from the server.
- Step 4: The server receives the symmetric key from the sensor node and then sends the confirm message which encrypted with the symmetric key just received to tell the sensor node that they have reached a consensus symmetric.
- Step 5: The next series communication between the sensor node and the server will encrypt the data with a symmetric key for a few times. For example, the symmetric key will be used for 100 times encryption. Then, the key will become invalid. They need to negotiate a new key from Step 1.

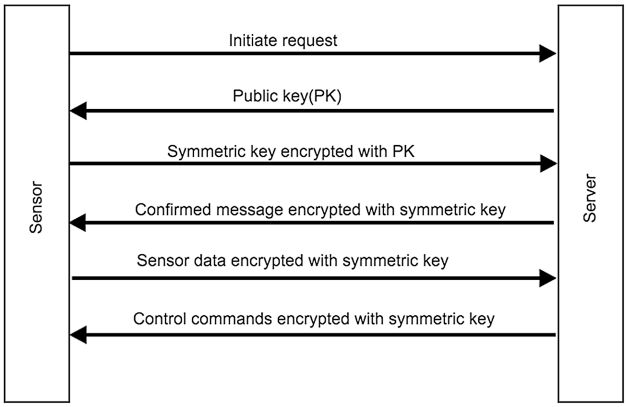


Fig. 4. The main encryption communication flow

In this framework, it uses RSA algorithm to achieve the goal of the distribution of key. In order to reduce the complexity of the encryption in the sensor nodes, we adjust the RSA algorithm, which makes the generated public/private key's length to be relatively short so that the sensor nodes can encrypt its randomly symmetric key in a relatively short time. The security of RSA is determined by the size of product of the two prime numbers chosen. In modern RSA encryption system, the product chooses 1024 bit usually. As the keys generation algorithm is complex, the server is responsible for the generation of the key pairs and sends the public key to the sensor nodes. Then the sensor node encrypts its symmetric key generated by itself with the public key after receiving the key from the server. After the encryption, the encrypted symmetric key will be sent to the server. The server will decrypt the encrypted message with the corresponding private key to get the symmetric key. The third part cannot decrypt the message because they don't know the private key. They have no necessary to try to crack the message because the protected symmetric key will only be used for only a few times and it will be changed randomly. What's more, the part of messages it encrypted have no too much value. It is not worth the cost to break the messages. In fact, the most of the messages are the similar sensor data. However, what it needs to protected with so much cost is the control data. We don't want the third part to know how we control the sensor nodes.

So here comes the question: how the nodes know that the message comes from the server instead of the fake? In order to fix the problem, we add a little more message during the step 2. This message is like a certificate which proves this message is from the server that we want to communicate with. The additional certificate will be encrypted with the server's private key. The content of the certificate will be previously stored in the sensor nodes' firmware. The sensor node will decrypt the encrypted certificate information with it received server's public key. Then the sensor node will compare the decrypted certificate information with the pre-stored one in the firmware. If they are same, the node can believe that the message is from the specific server rather than forging. In usually, the certificate should be download from the CA, but in this situation, we only need to ensure that the content of the certificate is not known by the

third party to ensure that the encrypted information is not being forged. Fortunately, the valid information is stored in the firmware where the third part cannot get it.

This process is complex, but it is also very important and necessary. Because if the messages transmitted through the network is cracked by some hackers, they can control the sensor node easily, even if control the mechanical equipment to make the plantings in a mess. But we can't guarantee that this situation will not happen. What we do is to reduce probability of the worst situation to be minimum.

The compute pressure of the whole process for the sensor node is not large. Because the public/private keys generation is processed in the server while the sensor node is only responsible for the encryption and decryption during the key distribution process and server authentication process. In the following data transmission, all the messages are encrypted with the symmetric keys where the compute pressure is not too heavy so that the sensor node is able to bear.

Distributive Optimization. It's easy to count the length of each message which the sensor nodes send to the server is 41 bytes. And the frequency of sending is once per second. Therefore, each sensor node will generate 144 kb data per hour and 3.38 MB every day. If the amount of equipment is about 1000, they will generate about 3.3 GB data which should be storage. Thus, it's necessary to take some solutions to storage the so much data in a proper way. At the same time, with the increase in the number of devices, the concurrency will be an inevitable demand when the number of request at one time is larger than thousands.

In order to solve the storage problem, we use the MySQL cluster as the main solution. We use multi MySQL server as the system's storage devices. Although the MySQL is a traditional database, but it has giant storage capacity if they can be organized as a cluster. And the cluster's query performance will be higher than the popular NoSQL databases in big data field, such as MongoDB, SequoiaDB, HBase, etc. [4] On the other hand, we use the load balance technology in the fourth layer (IP +Port) to solve the high concurrency problem to reduce the pressure of each server.

RESTful APIs for Mobile Apps Development. In order to let the farmers can control their equipment and adjust parameters in anywhere and anytime, the framework offers the mobile apps a set of RESRful APIs to invoke [5]. With the same RESRful APIs, whether Android apps or iOS apps, they can invoke them easily and quickly for all day long.

6 Conclusion

In this paper, we propose a TCP/IP network based IoT framework which had been applied into a practical project. It combines the mature and cheap electronic modules to construct a low-cost hardware which can sense the environment data in the fields and sent it to the server. What's more, it can receive the commands according the presetting from the remote server to control the related irrigation equipment through a safe way where security problem is a key point we considered in the framework. At the same time, it adopts the method of distributed database and the load balance technology to solve the problems of large data storage and high concurrency. In addition, it offers a set of RESTful APIs to the mobile apps to realize the mobile controlling.

Acknowledgement. This work is supported by the NSFC (61300238, 61300237,61232016, 1405254, 61373133), Marie Curie Fellowship (701697-CAR-MSCA-IFEF-ST), Basic Research Programs (Natural Science Foundation) of Jiangsu Province (BK20131004), Scientific Support Program of Jiangsu Province (BE2012473) and the PAPD fund.

References

1. Hong, S., Kim, D., Ha, M., Bae, S., Park, S.J., Jung, W., Kim, J.: SNAIL: an IP-based wireless sensor network approach to the internet of things. IEEE Wirel. Commun. **17**(6), 34–42 (2010)
2. Suo, H., Wan, J., Zou, C., Liu, J.: Security in the internet of things: a review. In: 2012 International Conference on Computer Science and Electronics Engineering (ICCSEE), vol. 3, pp. 648–651 (2012)
3. Pereira, P.P., Eliasson, J., Kyusakov, R., Delsing, J., Raayatinezhad, A., Johansson, M.: Enabling cloud connectivity for mobile internet of things applications. In: 2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE), pp. 518–526 (2013)
4. Botta, A., de Donato, W., Persico, V., Pescapé, A.: On the integration of cloud computing and internet of things. In: 2014 International Conference on Future Internet of Things and Cloud (FiCloud), pp. 23–30 (2014)
5. Kovatsch, M., Lanter, M., Duquennoy, S.: Actinium: a RESTful runtime container for scriptable internet of things applications. In: 2012 3rd International Conference on the Internet of Things (IOT), pp. 135–142 (2012)

Localization Technology in Wireless Sensor Networks Using RSSI and LQI: A Survey

Sai Ji[1,2(✉)], Dengzhi Liu[1,2], and Jian Shen[1,2]

[1] Jiangsu Engineering Center of Network Monitoring, Nanjing University of Information Science and Technology, Nanjing, China
jisai@nuist.edu.cn
[2] School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China

Abstract. For Wireless Sensor Networks (WSN), low-cost precise localization is the most essential requirement. Localization techniques based on RSSI is cost effective when be in comparison with TDOA, TOA and AOA. Because it doesn't need any extra power, hardware or bandwidth. In this paper, we simply introduce some related theory and techniques, such as TDOA, TOA and AOA in Wireless Sensor Networks. Localization error can be declined by observing both RSSI and LQI at the same time. We survey a dynamic distance estimation method based on RSSI and LQI, and present a comparison of some algorithms based on the theory. By analyzing the model of radio wave propagation loss and empirical data from real measurement, the method is to use discrete linear lines to approximate the real attenuation of RSSI and LQI.

Keywords: RSSI · LQI · Attenuation curve · Approximation · Linear lines

1 Introduction

After internet becomes universal around the world, Wireless Sensor Networks is one of the most influential information technologies which can deeply change the way of people's life. People can directly perceive the objective world through wireless sensor networks, as an extension to the existing Internet and greatly improve the ability of the human cognitive world. There isn't a uniform standard of localization in Wireless Sensor Networks. In this paper, we categorize them as range-based or range-free. In this kind of categorization, the characteristic of localization techniques is somehow indicated. Range-based techniques include use of radio, ultrasound or passive infrared. Due to no extra hardware, power or bandwidth, be utilized widely, Received Signal Strength Indicator (RSSI) becomes an ideal localization method. When communicating between nodes, localization is accomplished by measuring the received RF signal strength. Link Quality Indicator (LQI) connected with the radio gives data about the quality of the received signal and is strongly correlated with RSSI.

In this paper, we simply introduce some earlier localization techniques in Wireless Sensor Networks, such as TDOA, TOA, AOA. A brief comparison is showed following it. Section 2 is used to do so. The rest of this paper is organized as follows. In Sect. 3, we introduce RSSI and LQI, then we analyze why to combine them together.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_53

The comparison between three dynamic optimal processing algorithms is also presented in this Section. Finally, the conclusions of this paper are covered in Sect. 4.

2 Related Technologies and Theory

2.1 Received Signal Strength Indicator (RSSI)

RF signal is mainly used in RSSI. Receiving the signal from other transmitting nodes and measuring the received power, at the same time receiving point also calculate the values of the distance between two points according to the signal propagation model. Signal propagation model can be not only an experience model but also a theoretical mode. This way of measuring distance is easy to be affected by environment, and the nodes' ability to wireless communication is limited. Thus, RSSI is a kind of low precise and efficiency technique. Propagation model has a little complexity and inaccuracy. Therefore, the results are crude, even possibly to make ±50 % distance error. But it doesn't need any extra power, hardware or bandwidth. It's low-cost localization technique, if we can diminish the distance error, RSSI must be the most efficient localization technique.

2.2 Time of Arrival (TOA)

The technique calculates the distance by measuring the transmittal time of the signal between two nodes. TOA is mainly utilized in GPS systems. GPS systems require high energy consuming and expensive electronic equipment to keep the satellites in sync. But the sensor in the size of the hardware, energy consumption, the price limit, TOA technology is difficult to apply to Wireless Sensor Networks. Some papers present the simple use in localization based on TOA. As the Fig. 1 shown, pseudo noise sequence is used as the acoustic signal in the positioning part of the node. Obtaining the distance between nodes based on the propagation time of the acoustic wave. It consists of loudspeaker module, radio module, microphone module and CPU module.

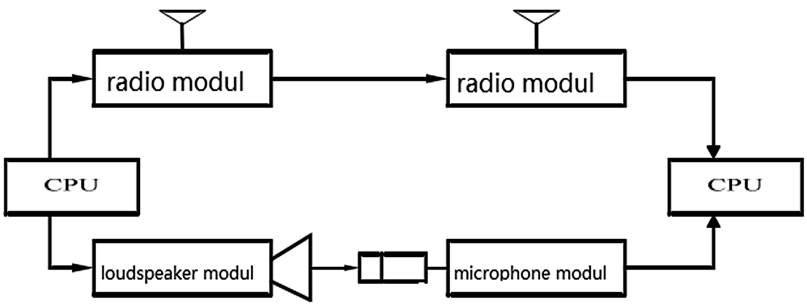


Fig. 1. TOA distance measure

2.3 Time of Difference on Arrival (TDOA)

TDOA technology is generally installed on the node RF and ultrasonic transceiver. Transceiver sends two signals at the same time, the use of electromagnetic waves and sound waves in the air propagation speed difference in the receiver transceiver to record the time difference between the two kinds of signals, and then turn the time into a distance. The transmitting node simultaneously launches the ultrasonic signal and the radio frequency signal, while the receiving node records the time of the arrival of the two kinds of signals T2, T1. As Known ultrasonic and radio frequency signal transmission speed is C2, C1, So the distance between two nodes is (T1 − T2) * S, specifically S = C1 * C2/(C1 − C2), as Fig. 2 shown.

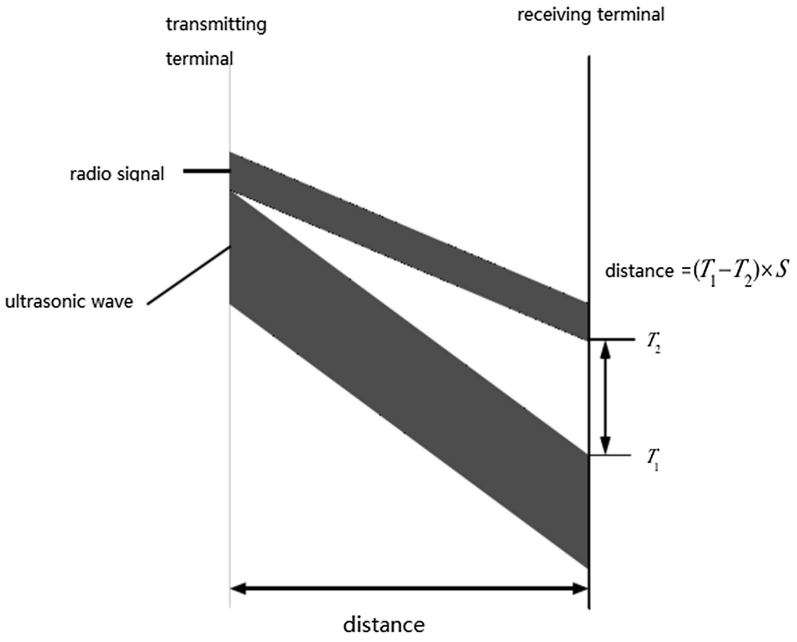


Fig. 2. TDOA localization theory

2.4 Angle of Arrival (AOA)

AOA ranging technology can estimate the direction of node transmission signal through a number of receiver or antenna array. The MIT Cricket Compass and other projects on the use of multi-receiver based on the AOA solution, the distance measurement system can be in the 40° angle to determine the direction of the error of 5°. AOA technology is easy to be affected by the external environment, and it may not be applied to sensor nodes in the size and power consumption of hardware. As Fig. 3 shows, the receiving node via a microphone array, perceived emission node signal direction of arrival, the following three stages on AOA determination of azimuth angle and position of the specific process. The first stage: measurement between neighboring

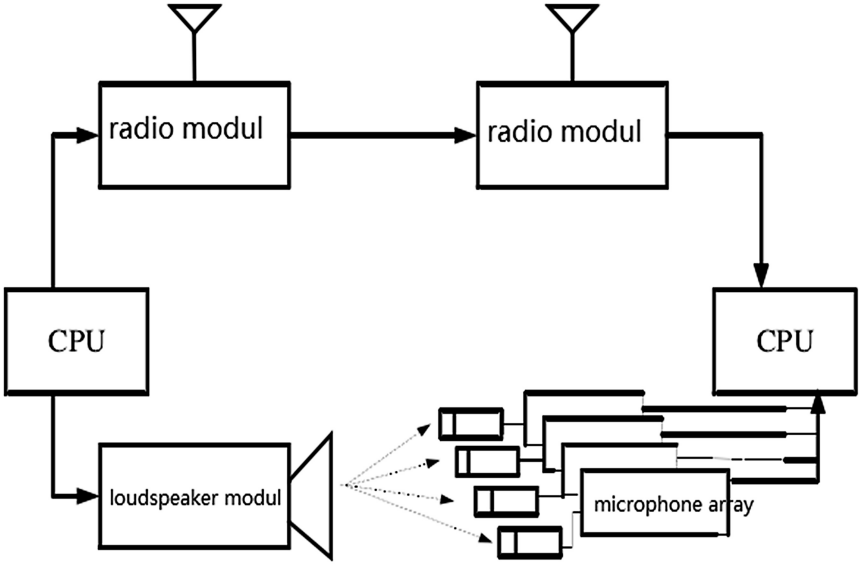


Fig. 3. AOA localization

nodes and azimuth angle; the second stage: relative measurement of anchor nodes azimuth angle; the third stage: the measured range information of the nodes compute node self-location.

Table 1. Characteristic of the four techniques

Category	Effective distance	Capacity of resisting disturbance	Measuring error	Extra hardware	Measuring cost
RSSI	Long	Multipath fading, subject to electromagnetic interference, be able to pass through barrier	Much	No	Little
TOA	Shorter	Affected by multi coating, reverberation effect, cannot pass through barrier	Less	Yes	Much
TDOA	Short	Affected by multi coating, reverberation effect, cannot pass through barrier	Little	Yes	Much
AOA	Short	Influenced by environment, cannot pass through barrier	Little	Yes	Much

2.5 A Brief Comparison

We summarize the four techniques and present a comparison Table 1, as introducing their characteristic in capacity of resisting disturbance, measuring cost, measuring error, extra hardware and effective distance.

3 Dynamic Distance Estimation Method Based on RSSI and LQI

3.1 Introduction of RSSI and LQI

RSSI is the difference between the received signal strength and the optimal received power level. Link quality indicator (LQI) is the characterization of the energy and quality of the received data frame. The value of LQI is based on signal strength and detected SNR (signal-noise ratio), and be calculated by MAC (media access control) level and provided up to the superior level. It's generally related to the probability of receiving a data frame correctly. The RSSI value and LQI value can be obtained when each data frame is received by the ZigBee transceiver module, which can reflect the change of signal intensity and the change of the interference on time. The dynamic range of LQI is larger than that of RSSI, with higher resolution. To integrate LQI value into the distance estimation algorithm, it's able to improve the localization technique based on RSSI.

3.2 Sectionalized Approximation of RSSI and LQI Decline Curves

Building RSSI and LQI decline curve is to find a simple and effective RSSI and LQI decline curve with changes in distance, and use mathematical formula to simulate decline curve, as the estimation of node distance basis. The actual application environment is much more complicated than the ideal environment, there is a multipath interference, diffraction, obstacles and other uncertain factors, for a specific environment, according to the principle of model and actual empirical data to establish empirical formula is the most effective method.

3.2.1 Classical Signal Decline Model

Path consumption model of radio propagation in free space is presented at formula (1):

$$\mathrm{PL(d)} = \mathrm{PL0} + 10\ \mathrm{nlg(d/d0)} + \mathrm{X}\sigma \tag{1}$$

d: distance between transmitter and receiver (km); f: frequency of radio transmission (MHz); n: path decline factor, generally be 2–5.

In the practical application environment, due to multipath, diffraction, obstacle factors, radio propagation path consumption and theoretical value compared with some of those changes, often using the lognormal distribution model formula (2):

$$\mathrm{PL(d)} = \mathrm{PL0} + 10\ \mathrm{nlg(d/d0)} + \mathrm{X}\sigma \tag{2}$$

PL: path consumption after d (dBm): d0: reference distance, generally be 1 m;

Xσ: Gauss distribution random variable with average value of 0, standard deviation is 4–10; PL (d0): path consumption after d0.

According to formula (1) and (2):

$$\mathrm{Pr(d)} = \mathrm{Pt} - \mathrm{PL(d)} \tag{3}$$

Pr(d): receiving signal strength (dBm).
Basing on the theory IEEE 802.15.4 give a formula (4)

$$\mathrm{RSSI(d)} = \begin{cases} p_t - 40.2 - 10 \times 2 \times \lg d, d \leq 8\mathrm{m} \\ p_t - 58.5 - 10 \times 3.3 \times \lg d, d > 8\mathrm{m} \end{cases} \tag{4}$$

Figure 4 shows the relationship curve between received signal RSSI and d in the model. From the channel model, it can be seen that the relationship between the attenuation of the signal and the distance to the exponential decay, and when the distance between the nodes is close, the path attenuation factor K is smaller (k = 2). When the distance increases, the attenuation increases (k = 3.3). Although IEEE802. 15. 4 simplify the channel model, the exponential calculation is relatively high to the hardware. From the signal attenuation characteristics can be seen in the distance, decline curve is steep, in the distance far away from the place, the curve becomes gentle. Therefore, it can be with a piecewise approximation method to fit the decay curve.

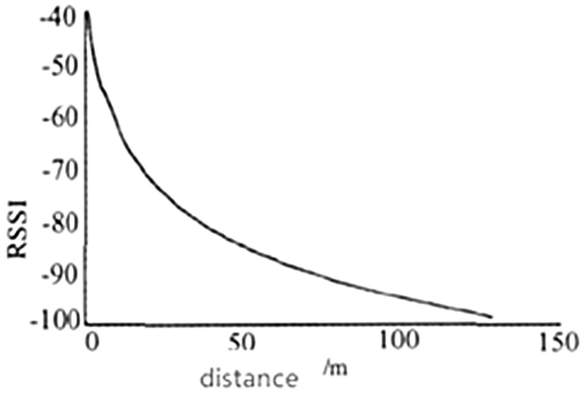


Fig. 4. RSSI-d

3.2.2 Establish Discrete Approximation Decline Curve of RSSI and LQI

The actual measurement of the RSSI and Q I L with the distance of the decline curve, it is as shown in Figs. 5 and 6. It can be observed from Fig. 5 that there are some irregular oscillation and fading in the actual decline curve. Signal intensity in the area less than 10 m better meet the 802. 15. 4 the signal decline curve, but when the distance increases, the RSSI decline curve relatively flat, significantly higher than the 802. 15. 4 the signal

attenuation curve. This may be due to the RSSI value is limited by the hardware conditions, its effective range is limited (measured under the conditions, generally in −40 ~ −90 dBm). Figure 6 shows the LQI with the distance of the fading situation. Its decline trend and RSSI are roughly the same: in the 10 m range of fluctuations is relatively small, with the increase of distance, the oscillation is also increased, there is a large irregular decline. In contrast, the effective range of Q I L value is much wider, between 170 and 40. Therefore, if the increase of Q I L value to carry out the distance estimation, can get a greater degree of linearity, so as to improve the positioning accuracy. According to the actual decline curve and the above analysis shown in Figs. 5 and 6, the linear approximation is carried out with two m as the boundary, and the RSSI and LQI fading curves of piecewise approximation are obtained as shown in Figs. 7 and 8.

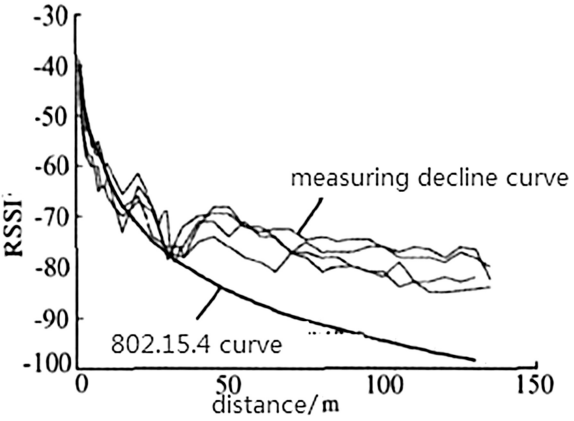


Fig. 5. RSSI measuring decline curve

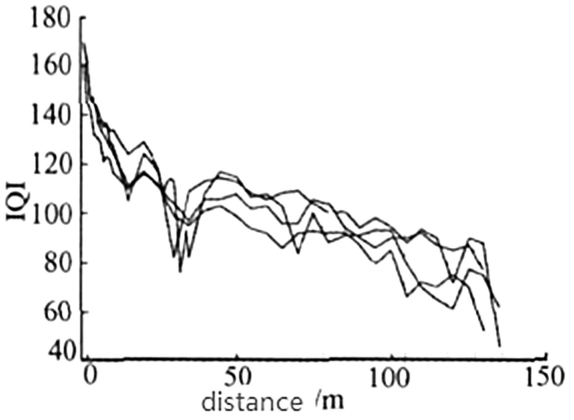


Fig. 6. LQI measuring decline curve

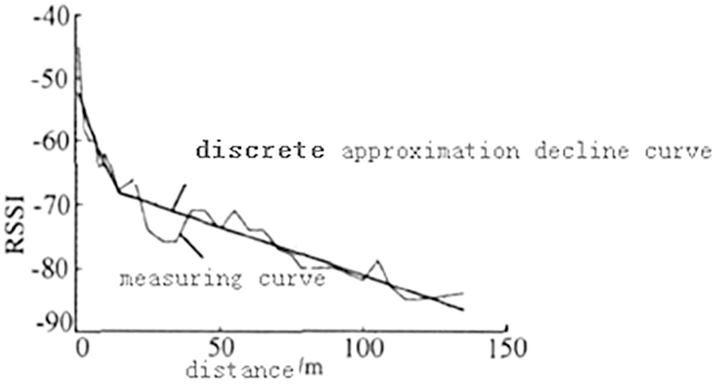


Fig. 7. Approximation decline curve

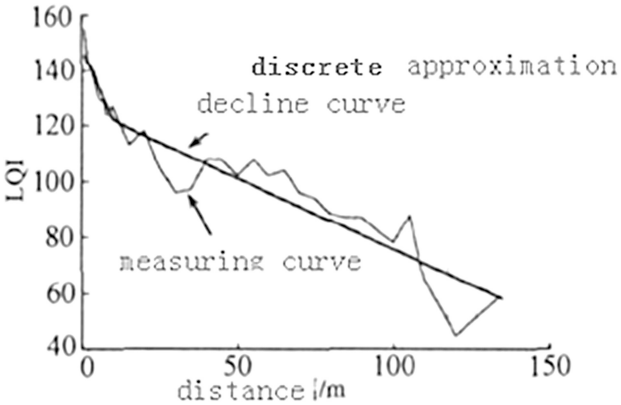


Fig. 8. Approximation decline curve

3.3 Dynamic Distance Estimation

From Figs. 7 and 8, we can see that the slope of the two LQI linear decline curve is higher than the linear approximation of RSSI decline curve, LQI can provide a higher resolution. In less than 10 m range, the actual LQI decline curve around the approximation curve fluctuations are slightly less than actual RSSI decline curve around the approximation curve fluctuations. It is relatively stable, but when the distance is expanding, LQI value fluctuation is more intense than the RSSI fluctuation, at this time, the distance which is calculated according to LQI approximation curve is very inaccurate, but RSSI is more stable. Therefore, if you can predict the distance of the range we need, when it's small, we choose LQI prior. When the distance is larger, we choose RSSI prior. So we can accomplish dynamic distance estimation.

The measured results show that the distance error calculated by LQI and RSSI is relatively small in the near range, and the difference between the two techniques is small. And when the distance is increased, due to the fluctuation of LQI caused by the

distance estimation also have a lot of error, which leads the difference between two techniques to be bigger. Thus, the distance difference can be used as the criterion for predicting the range of distance. By measured results can also be seen, in a position measured several times by a group of RSSI and LQI value if volatility is large, then the location of the real decline curve deviation curve approximation process also larger, using this to RSSI and LQI value in a certain range correction. We hypothesize that d1 is the estimating distance got by RSSI approximation decline curve, and d2 is the estimating distance got by LQI approximation decline curve, and d is the final estimation distance. There are three dynamic optimal methods, and a comparison is presented in next paragraph.

(1) Algorithm1: $d = a_1 d_1 + a_2 d_2$;

(2) Algorithm2: When $|d_1 - d_2| \leq t, d = d_2$;
Else $d = d_1$;

(3) Algorithm3: when $|d_1 - d_2| \leq t, d = d_2(\mathrm{LQI} \pm \Delta\mathrm{LQI})$;
Else $d = d_1(\mathrm{RSSI} \pm \Delta\mathrm{RSSI})$

3.4 Results

As Table 2 shown, row 1 and row 2 demonstrate that it's better to choose LQI, when the distance is short, and when distance is longer (>80 m), it's better to choose RSSI. Algorithm 1 is adjusted to the weight of d1 and d2, the performance of the algorithm is between the two, but the performance is improved in the whole range. Algorithm 2 can dynamically select RSSI or LQI to calculate the distance. Therefore, it is better than Algorithm 1 at all part. Because of including the correction of RSSI and LQI, Algorithm 3 is the best one in the three methods.

Table 2. Comparison of 3 dynamic optimal algorithm

Algorithm	0–10 m		10–80 m		>80 m		Entirety	
	Error	Variance	Error	Variance	Error	Variance	Error	Variance
RSSI	1.2267	2.374	15.173	391.60	6.4763	94.929	8.9067	179.45
LQI	0.8589	1.289	13.317	271.10	10.670	94.104	9.9311	214.50
Algorithm 1 $a_1 = 0.6$, $a_2 = 0.4$	1.0538	1.7786	33.584	322.50	6.7240	112.32	8.4819	162.05
Algorithm 2 $t = 8$	0.8589	1.289	12.043	268.06	6.4763	94.929	7.6178	135.30
Algorithm 3 $t = 8$, $j = 6$ $\frac{\Delta\mathrm{RSSI}}{2j}$ or $\frac{\Delta\mathrm{LQI}}{2j}$	0.7058	1.0136	12.1718	248.70	5.7846	86.720	7.3157	124.91

4 Conclusion

We've discussed some localization techniques in this paper. RSSI is utilized widely, because it's cost-effective characteristic. However, its performance is not as idea as we want. Hence, we combine it with LQI technique. Based on RSSI+LQI dynamic distance estimation, according to predicting how far the estimation distance is, we are able to dynamically select the best method to calculate distance. And we can do some correction to the algorithm by the fluctuation range of RSSI and LQI. In the future, we can subdivide the decline curve and increase liner approximation fragments. Therefore, we can improve the accuracy of distance estimation. We also will try our best to keep doing research, and strive to make more accomplishment.

Acknowledgement. This work is supported by the National Science Foundation of China under Grant No. 61300237, No. U1536206, No. U1405254, No. 61232016 and No. 61402234, the National Basic Research Program 973 under Grant No. 2011CB311808, the Natural Science Foundation of Jiangsu province under Grant No. BK2012461, the research fund from Jiangsu Technology & Engineering Center of Meteorological Sensor Network in NUIST under Grant No. KDXG1301, the research fund from Jiangsu Engineering Center of Network Monitoring in NUIST under Grant No. KJR1302, the research fund from Nanjing University of Information Science and Technology under Grant No. S8113003001, the 2013 Nanjing Project of Science and Technology Activities for Returning from Overseas, the 2015 Project of six personnel in Jiangsu Province under Grant No. R2015L06, the CICAEET fund, and the PAPD fund.

References

1. Bodhibrata, M., Sanat, S., Subrat, K.: Performance evaluation of localization techniques in wireless sensor networks using RSSI and LQI (2015)
2. Jieying, Z., Maohang, S., Xia, W.: Dynamic distance estimation method based on RSSI and LQI (2007)
3. Shen, J., Moh, S., Ilyong, C.: Comment: enhanced novel access control protocol over wireless sensor networks. IEEE Trans. Consum. Electron. **56**, 2019–2021 (2010)
4. Shen, J., Moh, S., Ilyong, C.: Enhanced secure sensor association and key management in wireless body area networks. J. Commun. Netw. **14**, 682–691 (2015). Minor revision
5. Dinca, S., Tudose, S.D.: RSSI-based localization in low-cost 2.4GHz wireless networks. Bucharest, Romania. Polytechnic University of Bucharest (2012)
6. Chengdong, W., Shifeng, C., Yunzhou, Z., Long, C., Hao, W.: A RSSI-based probabilistic distribution localization algorithm for wireless sensor network (2011)
7. Wang Y.: A research on the localization technology of wireless sensor networks (2007)
8. Ghelichi, A., Yelamarthi, K., Abdelgawad, A.: Target localization in wireless sensor network based on time difference of arrival (2013)
9. Gholoobi A., Stavrou, S.: Accelerating TOA/TDOA packet based localization methods (2014)
10. Shetty, A.: Weighted K-nearest neighbor algorithm as an object localization technique using passive RFID tags. M.S. Thesis, Rutgers University, New Brunswick, NJ (2010)

Energy-Balanced Unequal Clustering Routing Algorithm for Wireless Sensor Networks

Jin Wang[1], Yiquan Cao[1], Jiayi Cao[1], Huan Ji[1], and Xiaofeng Yu[2(✉)]

[1] School of Information Engineering, Yangzhou University, Yangzhou, China
[2] School of Business, Institute of Electronic Business (Yancheng), Nanjing University, Nanjing, China

Abstract. In wireless sensor networks (WSNs), the clustering routing technology can improve the scalability of the network. When the cluster head transmits data to the base station in a multi hop manner, the residual energy of cluster head and path condition are not considered. So it can reduce the lifetime of cluster head and seriously affect the network lifetime. We propose an energy-balanced unequal clustering routing algorithm for wireless sensor networks. Firstly, the non-uniform clustering method is applied to the network. Secondly when calculating the cluster radius, the residual energy of nodes, the density of nodes and the distances between the nodes and base station will be taken into account. Then, the algorithm establishes the shortest path tree to search the optimal multi-hop transmission paths to realize efficient data transmission from sensor nodes to base station. Simulation results demonstrate that the improved algorithm can efficiently decrease the dead speed of the nodes, balance the energy dissipation of all nodes, and prolong the network lifetime.

Keywords: Wireless sensor networks (WSNs) · Uneven clustering · Shortest routing tree · Network lifetime

1 Introduction

Wireless sensor networks (WSNs) are usually composed of large number of low-cost and tiny sensor nodes used for gathering information and then sending it in a multi-hop manner to some remote sink nodes. WSNs consist of a large number of small sensor nodes and accomplish distributed query tasks through wireless communication collaboration among them. The nodes can self-organize themselves to form a multi-hop network and transmit the data to certain sink node [1]. WSNs can be widely used to perform military tracking and surveillance, disaster relief, smart home, hazardous environmental exploration and health monitoring, etc.

The battery power is the most valuable resource of each tiny senor node in WSNs. This is because the battery power of sensor node is extremely limited, and it is usually difficult to replace them. One of the key issues in designing wireless sensor network communication routing protocol is to find a strategy to reduce the energy consumption and prolong the lifetime of the network. Therefore, it is important to design energy-balanced routing protocol for wireless sensor networks [2, 3].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_54

Clustering routing algorithm can reduce the energy consumption and improve the scalability of the network. LEACH protocol [4] is the first clustering routing protocol. LEACH protocol can effectively extend the network lifetime due to the reduction of the number of nodes to communicate directly with the base station. However, this protocol assumes that all nodes are homogeneous, therefore it cannot be applied to heterogeneous network environment. In most clustering routing protocol algorithms, collecting the data, data fusion and communication with the base station are usually completed alone by the cluster head nodes. Easily death of the sensors because the uneven energy consumption, severely reduce the life cycle of the network.

The remaining of this paper is organized as follows. Section 2 introduces some basic concepts and some clustering routing protocols for wireless sensor networks. Section 3 presents relevant system model. We propose our algorithm with detailed explanations in Sect. 4. Experiment results are shown in Sects. 5 and 6 concludes this paper.

2 Related Work

Regardless of the method of single hop or multi hop data transmission, there exist the problem of uneven energy consumption in uniform clustering method [5]. The communication between cluster head and sink node uses single hop mode, then, the energy of the node that is far from sink is quickly exhausted. Cluster head adopts multi hop forwarding can avoid this defect, but it will make the node closer to the sink node premature death.

For the first time, the idea of non-uniform clustering was first proposed by [6]. But the algorithm sets the cluster head having energy sufficient node in advance and there is no dynamic clustering. Chen et al. proposed an unequal cluster-based routing protocol in wireless sensor networks [7]. The algorithm partitions all nodes into clusters of unequal size, in which the clusters closer to the base station have smaller size. But the probability of cluster head selection is determined by the threshold value. So, it cannot guarantee that the selected cluster head is optimal.

Y. Zhou et al. proposed the improvement measures of cluster head selection and cluster head replacement, which can improve the performance of the network [8]. But each cluster head needs to broadcast information to whole network at the cluster formation stage, which will cause a great energy consumption. In [9], the authors proposed an uneven clustering routing algorithm for Wireless Sensor Networks (WSNs) based on ant colony optimization. The path factors such as bandwidth and transmission delay are considered when calculating the inter cluster path. However, the path pheromone is only partially updated, which is easy to cause uneven path load.

In view of the above shortcomings, this paper presents an energy-balanced unequal clustering routing protocol for wireless sensor networks. Firstly, the non-uniform clustering method is applied to the network. Secondly, the improved algorithm takes the residual energy, the density of the network nodes and the distance between the node and the base station into account in the selection of cluster head. Finally, we establish the shortest path tree to search the optimal multi-hop transmission paths to realize efficient data transmission from sensor nodes to base station.

3 System Model

3.1 Network Model

WSNs nodes are randomly distributed in a specific area and these nodes will eventually transmit the collected information to a base station. In this paper, we consider the random distribution of N nodes in a square area of $M \times M$. The following basic assumptions are made in this paper:

- The location of sensors node and the base station are fixed;
- Sensor nodes can communicate directly with the base station;
- Wireless links are bi-directional and symmetric;
- Sensor nodes can calculate the approximate distance of each other according to the value of the received signal strength;
- Sensor nodes can adjust the transmission power according to the distance of each other;

3.2 Energy Model

We use the similar energy model as that of [10]. Based on the distance between the source and target nodes, a free space or multi-path fading channel models are used. Each node will consume the following E_{tx} amount of energy to transmit a k bits message over distance d and the following E_{rx} amount of energy to receive the message.

$$E_{tx}(k,d) = \begin{cases} k \times E_{elec} + k \times \varepsilon_{fs} \times d^2, d < d_o \\ k \times E_{elec} + k \times \varepsilon_{mp} \times d^4, d \geq d_o \end{cases} \quad (1)$$

$$E_{rx}(k) = E_{elec}(k) = k \times E_{elec} \quad (2)$$

where E_{elec} means energy consumption per bit. ε_{fs} and ε_{mp} represent free space and multipath fading model respectively. d is transmission distance, and $d_o = \sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$ is a constant value.

4 The Proposed Algorithm

4.1 Selection of Cluster Head and Cluster Establishment

The energy-balanced unequal clustering routing algorithm for wireless sensor networks (EUCRP) adopts the idea of cluster head rotation, which can balance cluster member node energy consumption. The candidate cluster head probability is combined with the factor of nodes 'residual energy. Therefore, we redefine the threshold $T(n)$.

$$T(n) = \begin{cases} \frac{p}{1-p(r*\mathrm{mod}\frac{1}{p})} \times \frac{E_r}{E_{avg}}, & n \in G \\ 0, & n \notin G \end{cases} \tag{3}$$

where p is the desired percentage of cluster heads, r is the current number of rounds, G is the set of the nodes never become cluster head in recently $1/p$ round, E_r is residual energy of node, E_{avg} is the average energy of network.

Step 1: The base station broadcasts a clustering signal to the whole network, all nodes calculate the distance to the base station according to the received signal strength. Then, the competitive radius is calculated by the formula 4;

$$R_i = \left(1 - \omega \times (1 - \frac{1}{N(i)+1}) \times (\frac{d_{\max} - d(i, Bs)}{d_{\max} - d_{\min}})\right) \times \left(1 - \exp(-\frac{E_r}{E_o})\right) \times R_{\max} \tag{4}$$

where R_i is node i competing radius; $N(i)$ is the number of neighbors; $d(i, Bs)$ represents the distance from the node i to the base station; $d_{\max}$ is furthest distance far from the base station; $d_{\min}$ is the nearest distance far from the base station; ω is regulatory factor; $R_{\max}$ represents the maximum communication radius; E_r is residual energy of node; E_o is represents the initial energy of the node.

Step 2: Each node calculates their own threshold. If the threshold is less than $T(n)$, the nodes become temporary cluster head;

Step 3: Each node receives the election message from the other nodes and calculates the distance between the two nodes according to the received signal strength. The nodes that satisfy the formula 5 are added to the neighbor node table;

$$d_{ij} \leq \max(R_i, R_j) \tag{5}$$

where d_{ij} represents the distance between the node i and the node j.

Step 4: The candidate cluster heads calculate the waiting time t according to the formula 6. After waiting time t, they broadcast message that they become the cluster head. The other candidate cluster heads in its competitive range drop out of the race;

$$T = \lambda \times \mathrm{T}_{CH} \times \frac{E_o}{E_r} \times \frac{d_i}{d_{\max}} \tag{6}$$

where λ is random number between $0.9 \sim 1.0$, T_{CH} represent cluster head competition maximum time.

Step 5: After end of election, the nodes are not selected as cluster head node, they join the cluster which is nearest to the cluster heads.

4.2 Inter-cluster Multi-hop Routing

The inter cluster communication adopts multi hop routing, and the base station is the root of the tree. Each cluster is abstracted as a point. The adjacent clusters are connected together to construct a weighted connected graph $G = (V, E)$ that represent the wireless sensor networks, where V is the set of points (including the set of all cluster

heads and the base station), and E is the edge set. The transmission energy consumption of nodes, residual energy and the scale of cluster are factored into the weight of routing tree edge.

$$W(i,j) = \alpha \frac{E_{tx}(k, d_{ij})}{E_i} + \beta \frac{E_{rx}(k)}{E_j} \tag{7}$$

where $W(i,j)$ represents the weight of the edges of the adjacent cluster head node i and the node j in the graph; α, β represent random number between 0 $\sim$ 1.0; d_{ij} is the distance between node i and the node j.

Step 1: In a directed graph $G = (V, E)$, the cluster head node and base station in the set C are added to the set U, and the edge of the base station and the cluster head in the set C is added to the T;

Step 2: According to the received signal strength, they calculate the of each weight edge of the cluster head and the weight can be obtained by formula 7;

Step 3: From all of $u \in U$, $v \in V - U$ choose the minimum weight edge. The cluster head v will be added to the U, then the edge (U, V) is added to the set T;

Step 4: Repeat step 3 until $U = V$. A minimum spanning tree is constructed by all edges of the set T;

Step 5: According to the minimum spanning tree, the cluster head generates the shortest path routing, and transmits the data to the base station in a multi hop manner.

5 Performance Evaluation

To evaluate the proposed algorithm performance, it is compared with UCR algorithm [7]. Some of EUCRA algorithm parameters are set as follows: $\alpha = 0.4$, $\beta = 0.6$, $\omega = 0.5$, $p = 0.2$.

In the experiments, the whole network contains 200 sensor nodes in an $200 \times 200m^2$ area. All sensor nodes have the same initial energy of 0.5 J and they are randomly deployed inside the network. Some relevant network parameters are listed in Table 1.

Table 1. Network parameters

Parameter	Definition	Unit
E_{elec}	Energy consumption on circuit	50 nJ/bit
ε_{fs}	Free space model of transmitter amplifier	10 pJ/bit/m2
ε_{mp}	Multi-path model of transmitter amplifier	0.0013 pJ/bit/m4
l	Packet length	2000 bits
d_0	Distance threshold	$\sqrt{\varepsilon_{fs}/\varepsilon_{mp}}$
T_{CH}	The maximum competition time	0.5 s
$R_{\max}$	The maximum communication radius	50 m

Lifetime is an important metric to evaluate WSN algorithm performance. We evaluate the network lifetime using UCR algorithm and EUCRA algorithm respectively under the same network environment. As can be seen from Fig. 1, the number of nodes of the UCR algorithm starts to decrease near 680 rounds. However, the network using the proposed algorithm has the first dead node until about 980 rounds. As can be seen from the change of the survival number of node in network, the survival time of the nodes in the EUCRA algorithm is longer than UCR algorithm. So we can get a conclusion that EUCRA algorithm can efficiently improve the network performance for WSNs.

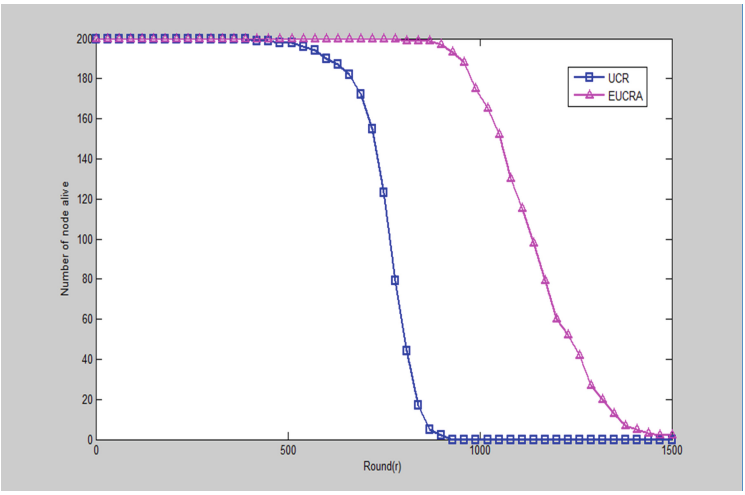


Fig. 1. Comparison of network lifetime

The amount of packets delivered by base station is studied in Fig. 2. It can be seen that EUCRA algorithm can deliver the amount of data packets about 1.5 times that of UCR algorithm at about 700 rounds. Our algorithm receives more data packets than the UCR algorithm. The main reason is that the proposed algorithm takes the residual energy of nodes, the density of nodes and nodes and the distances between the nodes and base station into account when calculating the cluster radius. The amount of packets are on rising phase at a slow rising speed between 900 round and 1200 round, because the appearance of death nodes and energy consumption of nodes, then the data packet received by the base station is gradually reduced. However, the network using the proposed algorithm continues sending data packets until about 1350 rounds. With the same initial energy, this work receives the amount of data packets than that of the UCR algorithm.

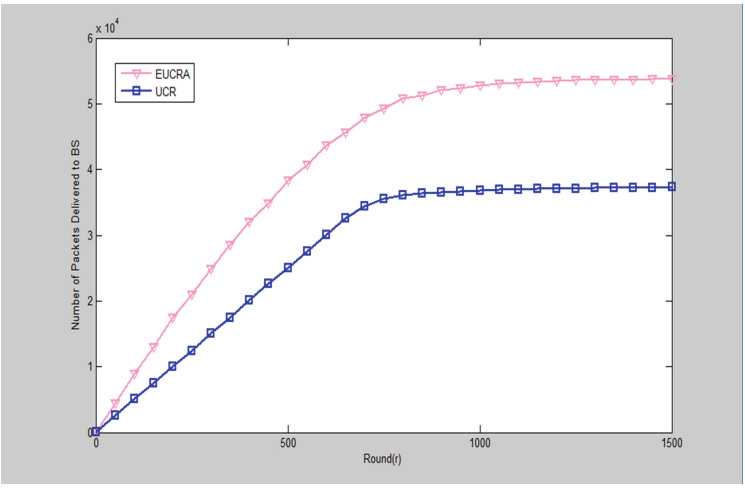


Fig. 2. Comparison of packet delivery

6 Conclusion

In this paper, we analyze the existing problems of clustering routing algorithm for wireless sensor networks and introduce the idea of non-uniform clustering. We propose an energy-balanced unequal clustering routing protocol for wireless sensor networks. In the selection of cluster head, the improved algorithm takes the residual energy, the density of the network nodes and the distance between the node and the base station into account. In the data transmission phase, the minimum spanning tree is built, and all cluster heads transfer the data through multi-hops to the base station. The experimental results show that the improved algorithm can effectively balance the network energy consumption, and prolong the network life cycle.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (61402234). Professor Xiaofeng Yu is the corresponding author.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., et al.: Wireless sensor networks: a survey. Comput. Netw. **38**, 393–422 (2002)
2. Wang, W., Zhang, S., Duan, G., et al.: Security in wireless sensor networks. IEEE Wirel. Commun. **15**(4), 60–66 (2008)
3. Tunca, C., Isik, S., Donmez, M.Y., et al.: Ring routing: an energy-efficient routing protocol for wireless sensor networks with a mobile sink. IEEE Trans. Mob. Comput. **PP**(99), 1–14 (2012)

4. Heinzelman, W.B., Chandrakasan, A.P., Balakrishnan, H.: An application-specific protocol architecture for wireless microsensor networks. IEEE Trans. Wireless Commun. **1**(4), 660–670 (2002)
5. Leu, J.S., Chiang, T.H., Yu, M.C., et al.: Energy efficient clustering scheme for prolonging the lifetime of wireless sensor network with isolated nodes. IEEE Commun. Lett. **19**(2), 259–262 (2015)
6. Soro, S., Heinzelman, W.B.: Prolonging the lifetime of wireless sensor networks via unequal clustering. In: Proceedings of the IEEE International Parallel and Distributed Processing Symposium. IEEE (2005)
7. Chen, G., Li, C., Ye, M., et al.: An unequal cluster-based routing protocol in wireless sensor networks. Wirel. Netw. **15**(2), 193–207 (2009)
8. Zhou, Y.C., Shi, R.H., Zhou, Y.Y.: Multi-hop routing algorithm based on uneven clustering for WSN. Appl. Res. Comput. **28**(2), 638–642 (2011)
9. Jiang, D., Liang, W.: Uneven clustering routing algorithm for Wireless Sensor Networks based on ant colony optimization. In: International Conference on Computer Research and Development, pp. 67–71 (2011)
10. Bajaber, F., Awan, I.: Adaptive decentralized re-clustering protocol for wireless sensor networks. J. Comput. Syst. Sci. **77**(2), 282–292 (2011)

The Agent Communication Simulation Based on the Ego State Model of Transactional Analysis

Mi-Sun Kim(✉), Green Bang, and Il-Ju Ko

Department of Media, Soongsil University,
Sangdo 1-dong, Dongjak-gu, Seoul, Korea
ronce81@gmail.com, banglgreen@gmail.com,
andy@ssu.ac.kr

Abstract. Inter-agent interaction plays an important role in cyberspace, which has recently been receiving attention as a space for social interaction. These interactions can occur between agents as well as between agents and people. However, for this, agents must be developed to the extent that they do not differ noticeably from human beings; that is, computers must be able to exhibit the same behaviors as humans. To achieve this, we developed a simulation in which human communication is replaced by a card game that corresponds to the ego state model of transactional analysis (TA). The agent communication module that we developed expresses agent-agent transactions based on the ego state model, as adapted to computer agents. The agents, the attacker and defender, continue the simulation with the cards that have messages and personalities. To win this simulation, the agents should find out their opponent's intentions and show a card that is suitable for their opponent's sample word or personality.

Keywords: Transactional analysis · Ego state · Card game · Agent communication

1 Introduction

Cyberspace has recently been receiving attention as a space for social interaction in many industries, including digital media. These interactions most often manifest in the fields of games, educational software, simulations and simulated scenario development, and film animation. Agent-agent interactions play a key role in these fields. However, for agents to interact naturally, they must be developed such that they do not differ noticeably from human beings. Thus, they should be capable of showing the same behavior that human beings do. Some researchers have managed to develop such agents, and adopted Transactional Analysis (TA), a theory based in psychology and psychiatry centered on personal growth, to develop a module capable of making agent–agent transactions more natural. In fact, researchers in other countries have defined a negotiation process for agents in which they act similarly to human beings using TA as a base [1], have developed a personality model for person–computer and person–agent communications [2–4], and studied the development of a personality model for

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_55

computer agents [5]. In other words, TA can be used as a personality model to express person–computer interactions, and even advanced transactions between agents.

The present study uses the ego state model of TA as the basis of a communication module for agents in cyberspace. In reality, people are normally able to see, hear, and feel others' words, facial expressions, tones, and behaviors to infer their internal states and maintain fluent communication. However, making agents grasp such informative data in cyberspace is far more difficult. Thus, in this study, we replaced normal human communicative information with cards, and simplified the simulation by applying the ego state model to both these cards and agents.

TA evolved as developer Dr. Eric Berne observed the behavioral changes occurring in patients when a new stimulus, such as a word, gesture, or sound, entered their focus [6]. TA posits that the major parts of the human personality are called the Parent (P), Adult (A), and Child (C) ego states. Berne defined an ego state as a "consistent pattern of feelings and experiences directly related to a corresponding consistent pattern of behavior." [7]. A transaction can be broken into two processes: a transactional stimulus and transactional response. These form the basic units of social discourse. TA defines three types of transaction, as follows. A complementary transaction is in which the ego state that is addressed is the ego state that responds. If picturing the three ego states with arrows or "vectors" to show the direction of each transaction, the vectors would be parallel. A crossed transaction is one in which the transactional vectors are not parallel; in other words, the ego state being addressed is not the ego state that responds. Finally, in an ulterior transaction two messages are being conveyed at the same time. One of these is an overt or social-level message, and the other a covert or psychological-level message.

2 Agent Communication Simulation

Before we explain the simulation, we must mention the words that were changed for the purpose of this study. The simulation takes the form of a game in which agents are either victorious or defeated. For the sake of convenience, the simulation is called *the game* hereafter. Furthermore, the term *ego state* becomes *personality* when applied to agents and cards; this is because *personality* is considered more suitable than *ego state* for the game format. This TA-based *game* is played in the form of agents persuading an opponent agent using cards with P, A, and C values. That is, the purpose of the *game* is to persuade an opponent by applying the ego-state model of TA to agents and cards. To win this *game*, the player's *personality* must not be exposed to their opponent; the player plays the *game* with cards corresponding to their opponent's intentions of wanting to understand the player's *personality*.

Table 1. Cards and P-A-C personalities

Major Intention	Major Personality	Major Action	Card Name
P→C	P	Judge, Sympathize	格言，鎮靜，大喝，訓戒，蔑視
A→A	A	Resolve problem	正義，道德，時間，思考，再考
C→P	C	Play together, Regret	反省，興奮，熱狂，詭辯，臆測

Each card has a different major personality—P, A, or C—and a major intention according to the dictionary definition of their name (Table 1). Furthermore, each card has a P-A-C value ranging between 5 % and 90 %; the sum of the P-A-C rates should be 100 % for each card. The major personality is decided using whichever personality has the largest figure value. For example, if one player has a card named "格言" and its major intention is P→C, then P refers to the personality of the "格言" card, while C means that the opponent has to play a C personality card in the next turn.

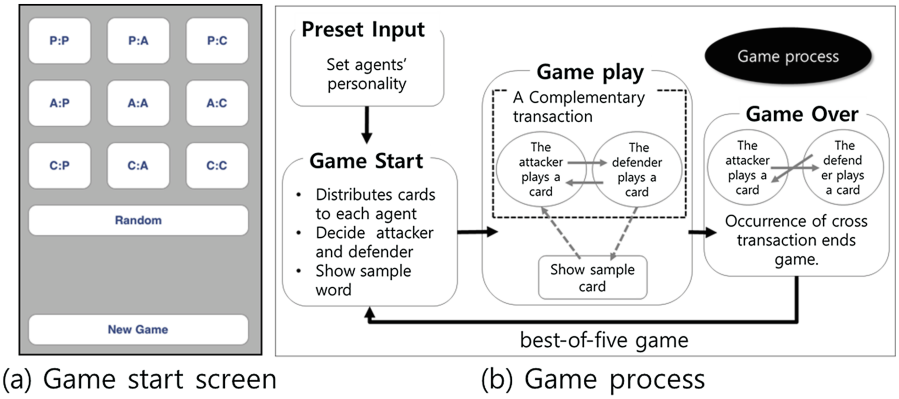


Fig. 1. Game design and game process

The game starts after agents' personalities are chosen. Agents' personalities were decided by an experimenter and the remaining process unfolded automatically (Fig. 1). When the game starts, each agent receives 10 cards and chooses the agent to attack. At this point, cards are divided based on the P-A-C personality of the agent. For example, when the agent has a personality where P is 20, A is 60, and C is 20, cards will be controlled by each personality with the probability of P, A, and C being 20 %, 60 %, and 20 %, respectively. After that, each agent receives a sample word to show their opponent and the game starts properly. Likewise, words are randomly chosen from the cards. The agent that has obtained the chance to attack shows a card first. The attacker can get the highest possible score when it shows a card that is most suitable to its personality. It can also score when it shows a card that is suitable for their opponent's sample word or personality. If it does not have a suitable card, it can discard any card that it currently holds, but it will then receive a low score or no score at all. The defender receives the highest possible score when it shows a card that is the most suitable to the attacker's card and its own personality. If it has a card that only suits either the attacker's card or its own personality, it can still show the card to the attacker. However, it will get a lower score.

For example, during the game, a situation may arise where the sample word is "格言," which has a P personality, the attacker's major personality is P, and the defender's major personality is C. In this case, the attacker shows a P→C personality card—which is suitable for the sample word—and the defender shows a C→P

personality card, which is suitable for the attacker's card and its own personality. As a result, both the attacker and defender receive points. In addition, by virtue of the successfully performed complementary transaction, the game continues. In contrast, if the attacker shows a card suitable to the sample word, but the defender shows a P→C card, which matches neither the attacker's card personality nor its own major personality, then the defender loses points. Moreover, as the defender showed a card that was not expected by the attacker, the game is over due to the crossed transaction. In other words, the more often a player succeeds in inducing a complementary transaction and persuading the opponent, the higher the likelihood they will win the game. Each game set consists of a total of five matches, and the agent who wins three matches first wins the game set. Whether the agent is victorious or defeated in the game is decided after the completion of one game set.

3 Result

Each game was played automatically except for the experimenter's choosing the agents' personalities before the game began. The game was played 10,000 times each for 9 situations consisting of P↔P, P↔A, P↔C, A↔P, A↔A, A↔C, C↔P, C↔A, and C↔C.

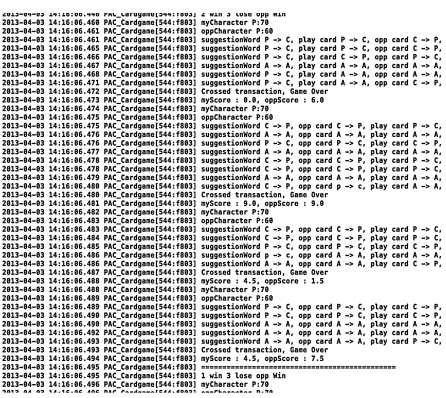


Fig. 2. A screen capture of the game process

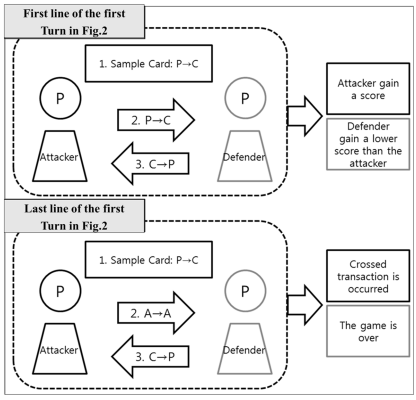


Fig. 3. Descriptions of Fig. 2

The game process was outputted as a text file (Fig. 2), displaying each game set (as three wins out of five matches), each agent's major personality and sample word, the cards shown by each agent, score per turn, and the victorious and defeated agents of each game set. In this game, we changed the topic into a sample word. The sample word is chosen among the cards that the agents exchange each other because it seems that the card can express the topic. The situation in Fig. 2 is one game set of P↔P. The first line in the figure; myCharacter P: 70 and oppCharacter P: 60 mean each agent's major personality. The meaning of myCharacter and oppCharacter in the myCharacter P: 70 and oppCharacter P: 60 is the name to distinguish each agent on the program.

Furthermore, P: 70 and P: 60 mean that each agent has 70 and 60 of the whole value, 100 as a P value. SuggestionWord means a sample word is outputted. The first suggestionWord is P→C and it presents the sample word's major personality is P and its major intention is C. However, there is P→C on the sample word, it is not considered its major intention (C), it is only used its major personality (P) in the game. In other words, the attacker can get a score when it shows P card that is suitable to the sample word's personality.

The term of play card shows first because myCharacter is the attacker in the first turn (Fig. 3). The card that myCharacter shows is suitable myCharacter's and sample word's major personality then myCharacter gets a score. OppCharacter shows C→P card according to P→C card's major intention that myCharacter showed, but the card's major personality (C) is differ from its major personality (P) so oppCharacter get a lower score than myCharacter. In the last of the first turn, a sample word is P→C but myCharacter doesn't have P→C card, it shows A→A card. And oppCharacter shows C→P card because it doesn't have A→A card. Thus, a crossed transaction is occurred and the game is over. The result of the game is that oppCharacter wins; myCharacter get 0 point and oppCharacter get 6 points.

The reason that the attackers get a lower score if they show cards that aren't suitable to the sample word's major personality because people are hard to gain the sympathy when they speak or act something that does not match the situations or the subjects in reality. Through the defenders show cards that match the attackers' card personality or not, it determines whether it is a complementary transaction or a crossed transaction. In reality, even if the sender speaks or acts something appropriately, the receiver doesn't humor, they can't keep their conversations. Thus, the term of finishing the conversation in reality is substituted for the game over in the game. Moreover each agent can get scores if it shows the same card as its major personality. This is involved with human personality, if they behave in accordance with other's mind but it is differ from their own personality, they may be stressed out. Therefore in the game, the agent loses scores when it shows a card that is not suitable for its own personality.

The game was played 4 turns in Fig. 2 and oppCharacter won 3 games. This means that oppCharacter shows more suitable cards to the sample word, its opponent's card or its own personality than myCharacter. This result presents that it is important that the agent is victorious or defeated is not decided which agent has a higher major personality value, but is decided how many cards, it shows, that is suitable for the sample word, its opponent's card or its own personality. The results were extracted by arranging the resulting values based on the winning rate of the attacker. In the 9 studied situations (P↔P, P↔A, P↔C, A↔P, A↔A, A↔C, C↔P, C↔A, and C↔C) the winning rate of the attacker was as follows: P↔P was 48 %, P↔A was 52 %, P↔C was 67 %, A↔P was 56 %, A↔A was 49 %, A↔C was 48 %, C↔P was 38 %, C↔A was 55 %, and C↔C was 54 %. Thus, most situations showed a similar winning rate for both the attacker and defender, except for the P↔C and C↔P situations, where the attacker's winning rate was 67 % and 38 %, respectively. This can be considered consistent with the prediction made in Sect. 2 on the analysis of TA theory that the P personality would have a higher winning rate than C in P–C communication, while the other situations would have similar winning rates. This also means most situations would show a similar winning rate for both the attacker and defender if they play the

game without agents' personalities. Therefore, TA theory should be employed to produce the agents that are more similar to human beings.

4 Conclusion

In the present study, messages that typically occur during human communication were replaced with cards reflecting the ego state model of TA to express agent–agent transactions. The agents themselves, interacting in a simulation game, were also governed by the ego state model. In the simulation, agents transacted with each other by exchanging cards and the expected results values were calculated. Through this process, we sought to present a standardized agent communication module.

However, the cards replacing human messages were not sufficiently diverse in the simulation, and the agent's personalities can hardly be viewed as sufficiently representing real human personalities. Furthermore, the simulation may have been rather simple and artificial, because we applied only basic structural analysis to the model explaining P-A-C ego states. In a future study, therefore, we intend to enhance the accuracy of the simulation using Egogram—diagrams that can standardize human personalities through ego states—and providing more diverse card messages. Furthermore, we intended to implement a more sophisticated simulation by adding functional analysis, which could aid in subdividing the P-A-C ego states into more specific categories. We expect this study to aid in better expressing various inter-agent transactions in cyberspace. In addition, we hope it will contribute to the development of more natural human-agent communication.

Acknowledgments. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032783).

References

1. Williamson, A., Ward, R.: Emotion in interactive systems: applying transactional analysis. Pers. Technol. **3**(3), 123–131 (1999)
2. Fujita, H., Sawai, N., Hakura, J., Kurematsu, M.: An action decision model for emotions based on transactional analysis. In: The 8th WSEAS International Conference on Artificial Intelligence, Knowledge Engineering & Data Bases, pp. 79–88 (2009)
3. Minamikawa, A., Yokoyama, H.: Blogs tell what kind of personality you have: egogram estimation from Japanese weblogs. In: Conference on Computer Supported Cooperative Work, pp. 217–220 (2011)
4. Katre, D.S.: Applying transactional analysis (TA) and ego-states for HCI design in e-learning: a case study of Indian scripts typing tutors. Eleltech India, pp. 16–20 (2005)
5. Booth, L.: The radiographer-patient relationship: Enhancing understanding using a transactional analysis approach. Radiography **14**(4), 323–331 (2008)
6. Harris, T.: I'm OK-You're OK. HarperCollins (2011)
7. Berne, E.: Games People Play. Grove Press, New York (1964)

Design of Software Reliability Test Architecture for the Connected Car

Doo-soon Park and Seokhoon Kim(✉)

Department of Computer Software Engineering,
Soonchunhyang University, Asan, Republic of Korea
{parkds, seokhoon}@sch.ac.kr

Abstract. Recently, the connected car is one of the hot topic in related business and industry. Although there have been many technical problems up to now, the evolution of various technologies has been resolved them. However, the software reliability and its test method in the connected car fields is very inadequate, because it is in the infancy step. To solve this problem, we propose the software reliability test architecture which can be applied to the connected car software. The proposed software reliability test architecture is designed to satisfy to 6 features which are functionality, reliability, usability, efficiency, maintainability, and portability. Based on this architecture, the software reliability test for a connected car can be used to verify a software, which is adapted to the connected car. In addition, the proposed test architecture can be supported or utilized a formal test procedure to the software in the connected car.

Keywords: Connected car · Cooperative intelligent transportation system (C-ITS) · Software reliability · Software test architecture

1 Introduction

Nowadays, according to the diffusion of smartphone deployments, various industrial fields, which are based on the Information and Communications Technology (ICT), have been evolved. Among the various fields, automotive industry has been reached a crossroads for important turning point based on the ICT convergence. That is, it has been not one of the most inevitable trends in these business areas. Therefore, the connected car, which can be connected to the Internet via various networking equipment, has been embedded many electrical devices, and can support multiple functionalities to the users. The software reliability in the vehicle has been one of the main issue, because the trends have been dramatically changing.

Actually, the software reliability of the vehicle's device was not a critical issue so far, because there are very few the software components, which are adopted to the vehicle's devices. However, the software reliability of the vehicle's devices in the practical life is now one of the most important issue, according to the accelerating diffusion of converged vehicles, such as a connected car, smart car, electric car, and so on. Although there are some automotive electrical and electronic system and development standards including a software, it is still insufficient to measure and test a software reliability of vehicle's device. Due to the fact that the various manufacturers,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_56

which are related to the converged car, produce the software by using their own criteria, and test a software reliability by using their own scenario, because there are no certain standards, frameworks, and architectures yet.

Basically, a software development process can't be specified, because a connected car is produced by various manufacturers. It is almost same as the software development process in other business areas. However, the developed software and its reliability test process is not same as the other fields, because the software reliability is closely related with the human safety. To solve these problems, we propose a new software reliability test architecture. The proposed architecture is similar to the legacy software reliability test architecture, but it is supposed to be used for the connected car software. In addition, the proposed test architecture can be used to measure a software reliability of vehicle's device, and can be utilized to test various functionalities and functional items.

The rest of this paper is organized as follows. Sections 2 describes the related works such as software reliability and its standards. Section 3 discusses the proposed software reliability test architecture. Finally, we offer concluding remarks in Sect. 4.

2 Software Reliability Engineering

A software reliability engineering has various definitions by multiple organizations. However, the most general definition is the application of statistical techniques to data collected during system development and operation to specify, predict, estimate, and

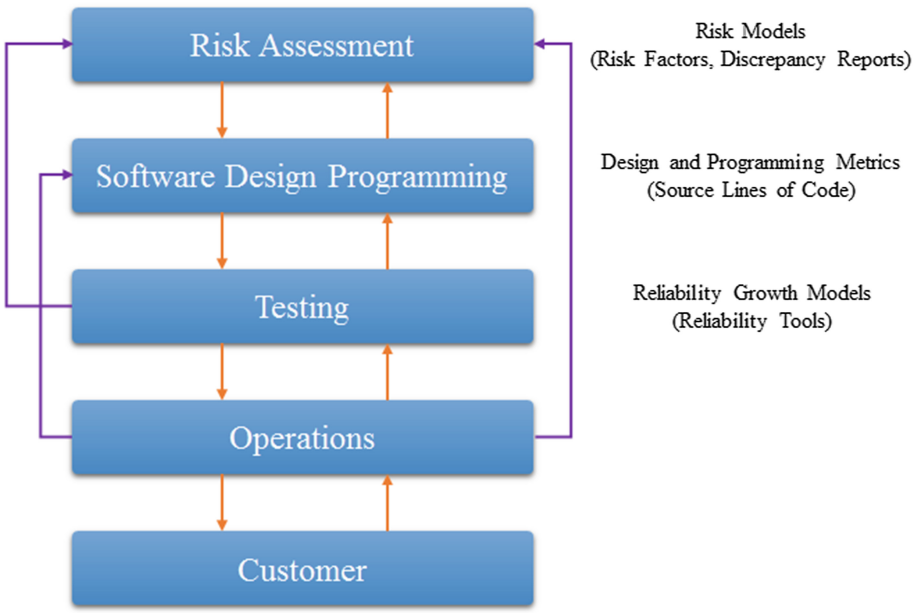


Fig. 1. Software reliability engineering process

assess the reliability of software-based systems by the American Institute of Aeronautics and Astronautics (AIAA) [1, 2].

Although there is a definition of software reliability as shown in Fig. 1, we can't directly apply it to the connected car's software, because we have to consider additional features for the connected car's software [3–6].

3 Software Reliability Test Architecture

Generally, a reference process model for developing products is V model in the ISO 26262 standards. As well known, ISO 26262 is consisted of ISO 26262-N: 2011 or ISO 26262-N: 2012. Among these standards, the part 4, part 5, part 6, and part 7 are related to the product developments, especially the software development process.

Table 1. The proposed architecture's features for software reliability test

Features	Descriptions
Functionality	Software product's ability when the software is operated in an explicit environment. The proposed architecture should be provided a concurrency and synchronization to the connected car's devices
Reliability	Software product's ability to maintain a performance in an explicit environment. The proposed architecture should be provided an emergency stop function to prevent a malfunction. In addition, the proposed architecture should support very flexible fault tolerance function based on the synchronization to the cloud networks
Usability	Software product's ability when the software can be easily utilized to the users. To do this, the proposed architecture should support various parameters, including default parameter values. Moreover, the architecture provides a various interface based on the HCI theory
Efficiency	Software product's ability to provide a required performance in an explicit environment. In the proposed architecture, the required performance can be varied by the resource amounts
Maintainability	Software product's ability to sustain a rebuilding or revision in an explicit environment. As same as the other architecture, the proposed architecture is depends on the hardware specifications, such as processor, memory, interface type, interface speed, and so on
Portability	Software product's ability to transit other environment, such as operating system. This feature in the proposed architecture is very important, because the portability is one of the main concern in the modern software development process

Although the defined software development process in the standards is very common aspects, it can't be suitable for the real software development process such as various embedded system software, including the connected car. Because of this, the proposed software reliability test architecture has the following features, as shown in Table 1.

To design the software reliability test architecture, we have to consider the following things, which are included the features which is described in Table 1. In this case, we assume that all software components in the connected car's device are series connections.

$$R_t = \prod_{1}^{n} R_i \tag{1}$$

where, R_t is entire system reliability, R_i is reliability of system i, n is number of components in series connections. To increase an entire system reliability, we have to guarantee each system reliability, as shown in Eq. 1. To do this, the proposed architecture is working with cloud system, which can support various software reliability features. Based on these scheme, the proposed architecture can be simple than the existing architecture. In addition, the proposed architecture can decrease number of components in series connections, because the proposed architecture is based on the cloud.

4 Conclusion

The software reliability test has been usually utilized to verify the various software functionalities in the computer system. However, in recent days, as a connected car gets the limelight, it has been a new understanding of the importance. In addition, the diffusion of IoT devices and services, and acceleration of convergence paradigm is one of the main reason. Moreover, the new applying fields of these technologies in the vehicle industry have been appeared in a connected car product.

Nonetheless, the software reliability and its test method hasn't been a main concern until now. This is why we propose the software reliability test architecture for a connected car. The proposed test architecture is a kind of test framework for a connected car's software. Based on the proposed test architecture, we can guarantee the software reliability of connected car. However, the proposed test architecture's performance can depend on the applicable test model. Therefore, we have to find the best solution and model in the future works.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP (Institute for Information & communications Technology Promotion), and this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2060035).

References

1. ISO, Road vehicles – Functional safety, ISO 26262 (2012)
2. IEEE, IEEE Recommended Practice on Software Reliability, IEEE Reliability Society, IEEE Std. 1633–2008 (2008)

3. ISO/IEC, Systems and software engineering — Software life cycle processes, ISO/IEC 12207:2008(E) (2008)
4. IEEE, IEEE Standard for Software Interface for Maintenance Information Collection and Analysis (SIMICA): Exchanging Test Results and Session Information via the eXtensible Markup Language (XML), IEEE Standards Coordinating Committee, IEEE Std. 1636.1–2013 (2013)
5. IEEE, IEEE Standard for Software Interface for Maintenance Information Collection and Analysis (SIMICA): Common Interface Elements, IEEE Standards Coordinating Committee, IEEE Std. 1636.99–2013 (2013)
6. IEEE, IEEE Standard for Automatic Test Markup Language (ATML) for Exchanging Automatic Test Equipment and Test Information via XML, IEEE Standards Coordinating Committee, IEEE Std. 1671–2010 (2010)

Network Activation Control According to Traffic Characteristics in Sensor Networks for IoT

Dae-Young Kim[1], Young-Sik Jeong[2], and Seokhoon Kim[3(✉)]

[1] Department of Software Engineering, Changshin University, Masanhoewon-gu, Republic of Korea
kimdy@cs.ac.kr
[2] Department of Multimedia Engineering, Dongguk University, Seoul, Republic of Korea
ysjeong@dongguk.edu
[3] Department of Computer Software Engineering, Soonchunhyang University, Asan, Republic of Korea
seokhoon@sch.ac.kr

Abstract. Sensor networks are an important technology for Internet of Things (IoT) connectivity. Sensor networks consists of numerous objects and the objects named sensor nodes collect information in access networks of IoT. They are transmitted the information to a network server which is placed in Cloud. In the sensor field (i.e., the access network of IoT), there are three kinds of traffic can be generated: event-driven, query-driven and periodic traffic. The traffic is classified and analyzed at the network server. The most of monitoring applications for IoT employs cluster-based network architecture. Each cluster gathers local sensing data and traffic characteristic of clusters is determined by the collected local data. This paper proposes network activation control of the clusters according to the traffic characteristic in order to minimize power consumption of sensor networks.

Keywords: Sensor network · Network activation control · Traffic characteristic · Internet of Things

1 Introduction

Internet of Things (IoT) provide intelligent services exploiting information from numerous objects. For the IoT services, it is important IoT connectivity to collect information. A sensor network is considered as an implementation technology of the IoT connectivity to gather and deliver the information, which is sensed by objects. It is applied to various IoT services such as industrial, military, environment, home and healthcare. That is, a sensor network has already become an access network technology for IoT [1–4].

A sensor network is based on power-aware systems. It consists of sensor nodes (i.e., objects) which include several capabilities such as sensing, computing, and communication. In addition, it employs RF for very low-power communication. The sensor

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_57

network aims to minimize energy consumption and maximize the network lifetime. Thus most of technologies for the sensor network focuses on the low-power computing [5–7]. There are many types of schemes to save consuming energy in aspect of sensor systems and data communication. However, effective method among them is controlling the network activation regionally. Because the most of sensor network systems construct clusters to collect regional information in monitoring applications, controlling the activation of the regional cluster is possible. That is, consuming energy in total network can be minimized through the power management for the network activation per cluster.

In addition, the sensor network has three kinds of traffic. They are event-driven traffic, query-driven traffic and periodic traffic. The event-driven traffic is generated when certain events are happened. The query-driven traffic is generated when a network server requests certain information to sensor nodes in the access network. The periodic traffic occurs with constant interval time. In the sensor network for IoT, the traffic is delivered to a network server via gateway nodes (i.e., sink nodes). The network server is placed in cloud and it processes the collected data from the access network. It can also manage the access network. Therefore, if the traffic characteristic of the collected data is considered at the network server, it can be expected the energy saving by controlling the network activation in a cluster. In this paper, we find the traffic characteristic in each cluster by statistical information of the collected data at the access network and then determine the network activation level according to the found traffic characteristic.

2 Network Model

In conventional sensor networks, there are two types of communication technologies. They are short range communication and long range communication. The short range communication employs ZigBee [8] based on IEEE 802.15.4 [9]. The long range communication includes LPWAN (Low Power Wide Area Network) such as LoRaWAN [10] or SigFox [11]. Both of the short range communication and the long range communication technologies aim for low cost and low-power communication with low data rate. For IoT services, the access network can have various wireless technologies such as ZigBee, Bluetooth, WiFi and LPWAN.

In this paper, a hierarchical network which consists of the short range communication and the long range communication is considered. As shown in Fig. 1, regional clusters are planned and each cluster employs the short range communication technologies (i.e., ZigBee, Bluetooth or WiFi) to collect local sensing data. Data, which is generated in the regional clusters, has one type among 3 types of traffic characteristic as mentioned earlier. Cluster heads deliver the data to network server using the long range communication technologies (i.e., LPWAN such as LoRaWAN and SigFox). The long range communication data is transmitted from cluster heads to a gateway node, which is interface between the access network and the IP network. Data at the gateway node is transmitted to the network server through the IP network. The network server controls

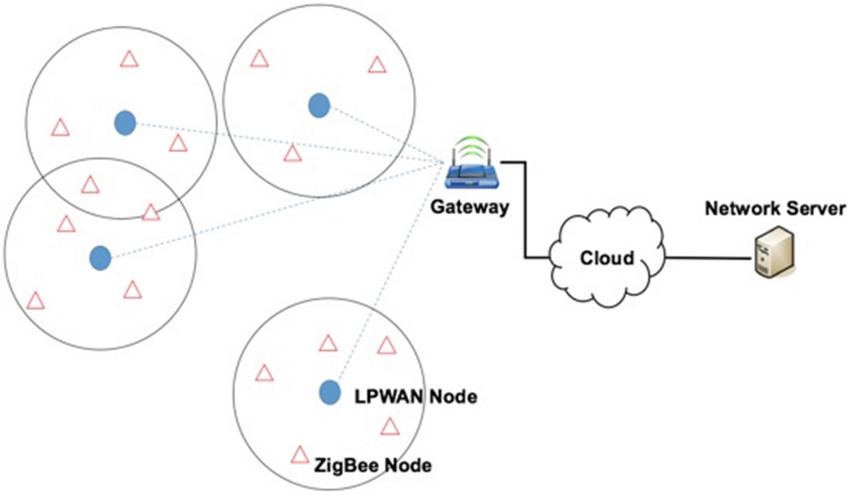


Fig. 1. Network architecture

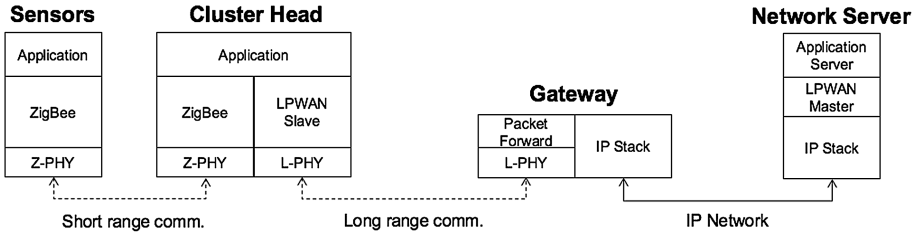


Fig. 2. Network model

the access network. In addition, it makes knowledges by the collected data and serves intelligent services by the knowledges (Fig. 2).

3 The Proposed Network Activation Control

3.1 Traffic Characteristic Determination for Clusters

As mentioned earlier, there are three types of traffic characteristic in the sensor networks for IoT services. In each cluster, similar data occurs and their types may be the same or very similar. In addition, in cluster-based networks, clusters can be managed by cluster heads. Thus, network operation policy according to traffic characteristic can be easily apply to single cluster. To apply different policies to each cluster, traffic characteristic of a cluster should be defined. In the proposed method, statistical histogram is exploited to find the traffic characteristic of a cluster. First of all, the network server manages the collected data from clusters. It employs window to make the histogram for the collected data of each cluster. Within the given window, the network

server makes the statistical histogram of the collected data type and then it defines the traffic characteristics of the cluster. According to the traffic type, the network server commands the activation policy for the cluster. Then, the cluster head operates its cluster according to the policy.

3.2 Network Activation Control

LPWAN end-devices (i.e., cluster-heads) construct star topology that centralized on the gateway. They can be synchronized by the gateway. The gateway sets activation period for the synchronized devices to control cluster activation. Then, it reports the period to the network server. The network server analyzes the histogram of each cluster and defines proper traffic characteristic for the clusters. After that, it determines activation policy of a cluster according to the traffic characteristic.

If a cluster has periodic traffic characteristic, it is not necessary that whole sensor nodes in a cluster activates all the time. Sensor nodes in a cluster periodically collect data at certain time. Thus, most of sensor nodes in a cluster can activate according to the data collecting activity. In this case, only few sensor nodes activates all the time to handle event-driven traffic. If a cluster has event-driven traffic characteristic, sensor nodes in a cluster should activate all the time in order to handle the event-driven data. If traffic characteristic of a cluster is query-driven, several times for handling query are set during the activation period. In case of the query-driven traffic characteristic, few sensor nodes activates all the time for event-driven data. Figure 3 represents the cluster activation time during the activation period.

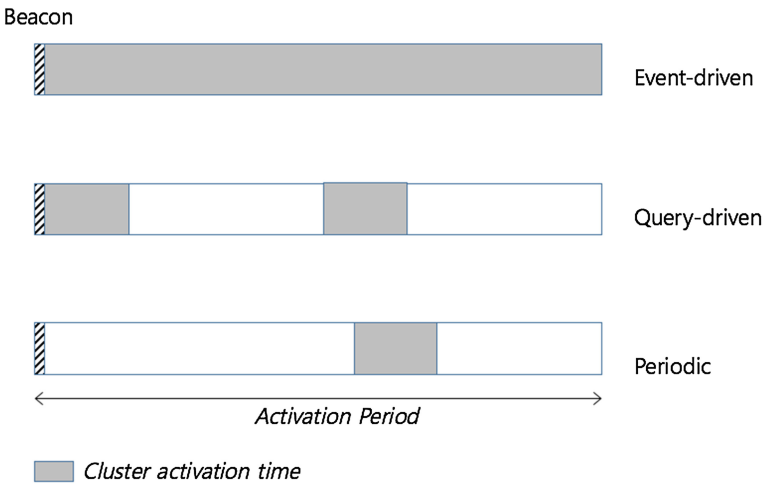


Fig. 3. Network activation according to traffic characteristic

4 Conclusion

Energy saving in the sensor networks is an important issue. Because objects named sensor nodes, which are battery-driven tiny devices, are composed of the network, low-power computing is necessary to maximize network lifetime. Although there are several methods to reduce consuming energy in the sensor nodes, most effective method is to control the network activation. To do this, the network is planned with clusters and the clusters are set to regional sensing area to collect local information. In this network, the proposed method exploited traffic characteristic of each cluster. The traffic characteristic was classified to three types: event-driven, query-driven and periodic. Each traffic characteristic requires different network activation time. Thus this paper exploited this property. According to the traffic characteristic, the activation of a cluster was determined.

Acknowledgement. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2014R1A1A2060035).

References

1. Atzori, L., Iera, A., Morabito, G.: The Internet of Things: a survey. Elsevier Comput. Netw. **54**(15), 2787–2805 (2010)
2. Gubbi, J., Buyya, R., Marusic, S., Palaniswami, M.: Internet of Things (IoT): a vision, architectural elements, and future directions. Elsevier Future Gener. Comput. Syst. **29**(7), 1645–1660 (2013)
3. Kim, S., Na, W.: Safe data transmission architecture based on cloud for Internet of Things. Wirel. Pers. Commun. **86**(1), 287–300 (2016)
4. Bandyopadhyay, D., Sen, J.: Internet of Things: applications and challenges in technology and standardization. Wirel. Pers. Commun. **55**(1), 49–69 (2011)
5. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirei, E.: A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
6. Culler, D., Estrin, D., Srivastava, M.: Guest editors introduction: overview of sensor networks. IEEE Comput. **37**(8), 41–49 (2004)
7. Kim, D.-Y., Jin, Z., Choi, J., Lee, B., Cho, J.: Transmission power control with the guaranteed communication reliability in WSN. Int. J. Distrib. Sens. Netw. (2015). ID632590
8. ZigBee Alliance. http://www.zigbee.org/
9. IEEE, IEEE standard for information technology—telecommunications and information exchange between systems—local and metropolitan area networks—specific requirements—part 15.4: wireless Medium Access Control (MAC) and Physical Layer (PHY) specifications for low-rate wireless personal area networks (LR-WPAN), IEEE Standard 802.15.4-2006 (2006)
10. LoRa Alliance. https://www.lora-alliance.org/
11. SigFox. http://www.sigfox.com/

Data Mining Techniques to Facilitate Digital Forensics Investigations

Erik Miranda Lopez[1], Yoon Ho Kim[2], and Jong Hyuk Park[1(✉)]

[1] Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul 139-743, Republic of Korea
{erik.miranda,jhparkl}@seoultech.ac.kr
[2] Department of Division of Computer and Media, Mokwon University, Daejeon 35349, Republic of Korea
yhkim@mokwon.ac.kr

Abstract. Digital forensics is an essential discipline for both law enforcement agencies and businesses. It makes possible to investigate electronic related crimes aka cybercrime such as fraud, industrial espionage and computer misuse. However, encryption, anti-forensic tools and the ever increasing amount of volume of data to analyse creates a wide range of challenges to overcome. Fortunately, other computer fields can be applied to overcome those challenges. This paper will explore some data mining techniques to address most common issues in Digital Forensics.

Keywords: Digital Forensics · Data mining · Deviation detection · Entity extraction · Classification

1 Introduction

Information Technology is being abused as an instrument for criminal activity and Digital Forensics (DF), also called Forensic Computing, is the use of computer technologies to investigate such offenses. DF investigations are not limited to law enforcement agencies though. In the private-sector, businesses use DF to investigate email abuse, employee terminations and industrial espionage cases [1].

Forensic computing is a relatively new discipline with still many challenges to overcome. Challenges range from lack of standards and tools, the never ending increase of data volume and the high complexity of data analysis in investigations.

Data mining, with its data analysis techniques to discover interesting patterns, can be used to aid DF practitioners. Data found on evidence can be analysed with statistical methods, machine learning and data mining algorithms. These data mining techniques are being applied to solve some of the main challenges found in forensic investigations.

The aim of this paper is to explore the challenges that DF is currently facing and discuss some data mining techniques that are used to facilitate forensic investigations with large datasets. The data mining techniques reviewed in this paper are deviation detection, entity extraction and classification.

There are a number of definitions for digital forensics. As defined by computerforensicsworld.com [2], DF is the discipline that involves the use of "*analytical and*

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_58

investigative techniques to identify, collect, examine and preserve evidence/information which is magnetically stored or encoded". McKemmish [3] presented DF as the "*process of identifying, preserving, analysing and presenting digital evidence in a manner that is legally acceptable*". US-CERT [4] has its own definition too: "*The discipline that combines elements of law and computer science to collect and analyse data from computer systems, networks, wireless communications, and storage devices in a way that is admissible as evidence in a court of law*". We can clearly see a set of keywords in these definitions: collect, preserve, analyse and admissible in court.

Forensic computing investigation takes place after an incident has occurred and it can assist in a wide range of cases:

- IP Theft
- Criminal Damage
- Industrial Espionage
- Fraud Investigations
- Corporate Policy Violation
- Child Pornography
- Computer Misuse
- "Defence-in-depth"

There are a wide range of challenges in DF, from a legal and administrative point of view; lack of standards, lack of international cooperation and 'law lag'. From the technical side; encryption, anti-forensic tools, data volume and new technologies to mention a few [5]. This paper focuses on the technical challenges posed by the ever increasing data volume.

2 Data Mining in Digital Forensics

2.1 Data Mining

Data mining is an interdisciplinary topic to analyse vast amounts of data and extract knowledge from it [6]. Simply put, data mining is the use of methods and algorithms to extract information from raw data. Analysing this tremendous volume of data is an important need to discover knowledge which may help in making important decisions. Use cases include retailers using the knowledge for optimising their inventory, pharmaceutical companies to explore drugs interactions and marketing industry to target potential consumers. As we'll see in the next section of the paper, data mining can also address some of the biggest challenges in forensic computing.

2.2 Data Mining Techniques to Aid Digital Forensics

- Deviation Detection aka anomaly detection or outlier detection refers to describing the most substantial changes from previously measured or normative values [6]. This data mining technique can be useful for investigating computer system intrusions, digital forgery [7] and crimes that require tracing unusual activities [8].

Another useful example where anomaly detection can be applied is in fraud detection. A forensic investigator may be able to detect suspicious or fraudulent activity for insurance claims, bankruptcy or tracing terrorism by examining financial records.

- Entity Extraction is the technique that identifies specific patterns from big datasets [9].
  It is a valuable technique when investigator are looking for personal information or any data that can help identifying a person, object or action. For instance, when investigating a server compromise, entity extraction will help finding login ID, IP, time and other relevant information from the logs [10]. This allows a quick and efficient extraction of information from large distributed databases or logs.
- Classification finds common properties and organises them into classes.
  A common used case is tracing spam. This method aids investigators identifying the source of spam by analysing the sender's written patters and structural features [11]. The main drawback of this technique is the need of large data samples to build a pattern [9] (Table 1).

Table 1. Data mining techniques in digital forensics

DM technique	DF task	Use cases
Deviation detection	Data analysis	Computer system intrusions. Fraud detection. Digital Forgery
Entity Extraction	Data analysis	Looking for specific information in large datasets
Classification	Data analysis	Spam tracing

As outlined, the ever increasing volume of data poses one of the biggest challenges in digital forensics today. Although this paper has only explored a few data mining techniques available, there's still more research to be done to address the issues created by the enormous amounts of data. As seen, data extraction techniques can greatly reduce the time needed in an investigation. Nevertheless, more work needs to be done to reduce the amounts of real-world data.

3 Conclusion

We have seen digital forensics as the discipline that collects, preserves and analyses evidence and it's used by both law enforcement agencies and businesses. We've also briefly described the challenges computer forensics is facing, particularly how to deal with large volumes of data and extract relevant information from it. In order to address this issue, we've explored some data mining techniques; deviation detection to investigate intrusions, digital forgery and fraud cases; entity extraction to extract valuable information from large datasets; and classification to structure written patters in order to trace the origin of emails.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP(Institute for Information & communications Technology Promotion).

References

1. Nelson, B., Phillips, A., Steuart, C.: Guide to computer forensics and investigations, Cengage Learning (2015)
2. Computer Forensics World (2016). http://www.computerforensicsworld.com/. Accessed 4 May 2016
3. McKemmish, R.: What Is Forensic Computing?. Australian Institute of Criminology, Canberra (1999)
4. US-CERT, Computer Forensics (2008). https://www.us-cert.gov/sites/default/files/publications/forensics.pdf. Accessed 14 May 2016
5. Mercuri, R.: Challenges in forensic computing. ACM **48**(12) (2015)
6. Han, J., Kamber, M., Pei, J.: Data mining: concepts and techniques. Elsevier (2011)
7. Mahdian, B., Saic, S.: Using noise inconsistencies for blind image forensics. Image Vis. Comput. **27**(10), 1497–1503 (2009)
8. Justickis, V.: Criminal datamining. Security Handbook of Electronic Security and Digital Forensics (2010)
9. Chen, H., Chung, W., Xu, J.J., Wang, G., Qin, Y., Chau, M.: Crime data mining: a general framework and some examples. Computer **37**(4), 50–56 (2004)
10. Sindhu, K.K., Meshram, B.B.: Digital forensics and cyber crime datamining. J. Inf. Secur. **3**(3), 196 (2012)
11. de Vel, O., et al.: Mining e-mail content for author identification forensics. SIGMOD Rec. **30**(4), 55–64 (2001)

Solving Standard Cell Placement Problem Using Discrete Firefly Algorithm: A Nature Inspired Approach

Pradip Kumar Sharma, Saurabh Singh, and Jong Hyuk Park(✉)

Department of Computer Science and Engineering,
Seoul National University of Science and Technology, Seoul, Korea
{pradip, singhl989, jhparkl}@seoultech.ac.kr

Abstract. Standard Cell Placement is a vital step in the digital circuit layout that allots positions for several circuit components within the chip's core region. Placement problem has drawn broad explores tending in the VLSI CAD domain. The purpose of this research is to present a discrete approach based on firefly algorithm for standard cell placement problem. The results show that the proposed DFCP approach are better than mete genetic approach for the placement instances of MCNC benchmark. The research testifies that the DFCP approach provides effective cell locations, giving a super-ordinate performance in terms of time and quality of the solution.

Keywords: VLSI · Firefly algorithm

1 Introduction

Nanotechnologies that actually exist today, the complexity of Very Large Scale Integration (VLSI) circuit design have breakneck. In fabrication technology, laying out of the circuit is a most challenging chore of a digital circuit design that directly touches on the performance, area, yield and reliability of the circuit. To reduce the layout design complexity, the problem is carved up into smaller sub-problems that are solved one by one afterward.

A congestion-oriented approach can hardly be successful without a proficient quality wire-length minimization engine. Design techniques allow a semi-regular layout in which all cells bear the same height and their width changes with their functionality. This layout can be carried out in a comparatively short time. In standard cell layout, one major target is to reduce the overall interconnect length by putting the cells optimally with respect to one other.

Given an electrical circuit consisting of cells, with predefined input and output terminals, interconnected in a predefined way. The Standard Cell Placement problem builds a provision indicating the positions of the cells so that the estimated length of the wire and the layout area is minimized data and other constraints are met. The inputs to the problem is the description of the cells with sizes and terminal locations, and the netlist defines the interconnections between cells. The output list contains a list of coordinates X and Y cells.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_59

The standard cell placement can be more formally stated as: Given a circuit represented by a hypergraph G(V, E), where the vertex set $V = \{v_1, v_2, \ldots, v_n\}$ denotes the set of cells to be placed and the edge set $E = \{e_1, e_2, \ldots, e_n\}$ denotes the set of nets, representing cells connections. Each cell i is allocated a position (x_i, y_i) on a rectangular area of the chip. A non-negative weight w_j is assigned to each edge e_j. Then using HPWL (i.e. Half Perimeter Wire Length), the total wirelength i.e. W(x,y) in below equation is calculated.

$$W(x, y) = \sum\nolimits_{e \in E} (\max|x_m - x_n| + \max|y_m - y_n|), \text{ Where } v_m, v_n \in e$$

The solution to the problem lies in packing all the cells into the design area and finding their optimal positions such that there is no overlap among cells and the total wirelength of the placed circuitry is minimized.

The idea of partition-driven placement initiated by Breuer [1] using the extension of Kernighan and Lin algorithm [2] by repeated graph bisections. The author observed better half-perimeter wirelength (HPWL) using alternating cut-line directions than using only horizontal cuts.

Jiang [3] developed a hybrid placer, NTUPlace2 that used analytical techniques to aid partitioning for standard-cell designs. The placer firstly positions the objects in a placement bin by an analytical technique to cut down quadratic wirelength and then partition it with hMetis partitioner.

Subbaraj P. [4] proposed a parallel genetic algorithm that reduces the phase calculation of the genetic algorithm using the parallel approach. The proposed methodology is verified in different circuits of IBM. In terms of quality of the solution and computing time, the proposed method is found better as compared to Fastplace, Dragon 2.2.3 and Qplace5.1.67.

Myung-Chul Kim [5] proposed a new global placer tool, SimPL. SimPL that utilizes a self-contained, flat, quadratic technique. It maintained lower-bound and upper-bound placement for converging to a final solution.

Recent metaheuristic algorithms have been advanced with an objective to achieve global search with three main goals: obtaining robust algorithms, solving large problems, and solving the problem faster.

At Cambridge University, the author Xin-Shw Yang is proposed the firefly algorithm (FA) [2, 6–12]. It is a new metaheuristic approach, which is inspired by the behavior of fireflies. It belongs with the nature-inspired algorithms, originally designed to solve continues optimization problem.

2 Proposed Methodology

The proposed methodology denotes each firefly is a solution. Our goal is to generate new solutions that might better firefly replacement of the previous poor solutions. The algorithm is inspired by the attractiveness and brightness of fireflies generally seen in the summer sky in several regions. These features help the fireflies to protect themselves from piranhas and attract their prey, making the fireflies move towards more lighting points to find a global best possible solution.

2.1 Firefly Solution

For the standard cell placement as shown in Fig. 1, a solution presents a permutation coding. A solution is denoted as a firefly. An array of element denotes a circuit's cell.

With the initial set of arbitrary population, the algorithm begins. A size of m × n matrix is defined to signify the position of the life of all fireflies. Whereas n is cells number and m is a solutions number.

8	1	2	7	4	9	5	3	10	6

Fig. 1. Firefly solution

2.2 Distance Between Fireflies

An different arcs number indicate the distance among the fireflies for a given set of solutions.

Considering different fireflies, the distance among them is calculated by

$$d = \left(\frac{R}{N}\right) * S$$

Between two different fireflies, if R represents the number of total unique arcs, N – is the total cells and S – represent success factor depending upon the number of cells.

Between the Firefly i and j, the number of different arcs is six. Such as shown in Fig. 2, six arcs 8-1, 2-7, 7-4, 9-5, 5-3, and 3-10 in Firefly i do not contains in Firefly j.

Firefly i

8	1	2	7	4	9	5	3	10	6

Firefly j

8	2	1	7	3	9	4	10	6	5

Fig. 2. Different arcs between fireflies i and j

2.3 Attractiveness

Each firefly has its attractiveness β described by monotonic decreasing function of the spaced between any two fireflies

$$\beta(r) = \beta_0 e^{-\gamma d^2}$$

Where r is the distance between two fireflies, γ is a light absorption coefficient and β_0 is the attractiveness at r = 0 (Table 1).

Table 1. The pseudocode of the proposed DFCP algorithm

```
Output:
Smallest Wirelength
Placements of all the cells in Netlist
Input:
   Objective function
   Benchmark data
Begin
   Firefly Solutions Xi, (i = 1, 2,… ,N) Initialize Arbitrarily
   Evaluate fitness of all solutions
   Take local_best_solution
   Loc_Gen := 1
   While Loc_Gen < Threshold_Value    do
      For i = 1 to Pop_size
         If (Fj >Find_Attractiveness(Fi))
            Then Move firefly i towards j in dim- dimension
         Else
            Move firefly i randomly in dim-dimension
         Endif
      Endfor
      Take the best local solutions
      Determine the local_best_solution from (dim*
Pop_size)+1
   Endwhile
```

3 Simulation Results

The DFCP algorithm is tested on reference circuits MCNC. We compared the obtained results with meta-genetic approach. For multiple instances of circuits of MCNC benchmark, the runtime and total wirelength of the proposed algorithm are presented in Table 2.

Table 2. Comparison of wirelength and CPU time of the DFCP algorithm and meta-genetic approach

Benchmark circuits	No. of cells	No. of nets	The DFCP algorithm		Meta-genetic algorithm	
			Avg. total wirelength	CPU time (in sec)	Avg. total wirelength	CPU time (in sec)
Fract	149	147	192	21	211	30
Primary1	833	904	3413	158	3785	214
Struct	1952	1920	325	120	337	142
Primary2	3014	3029	23632	237	24902	264
Industry1	3085	2594	13495	212	14021	230

As seen from the table, the Avg. wirelength as compared to the meta-genetic approach results from the proposed algorithm are obtained better.

4 Conclusion

This work has proposed a heuristic approach based meta swarm is proposed to solve the standard cell placement problem. The proposed algorithm includes a discrete approach firefly (DFCP) for better improvement of the solution. The algorithm is tested on various circuits Benchmark MCNC. Experimental results have shown that the proposed algorithm gives better results than the meta-genetic approach. Furthermore, the results highlight the flexibility of the proposed method for solving the NP-complete problem of the standard cell placement.

References

1. Breuer, M.A.: A class of min-cut placement algorithms. In: Proceedings of Design Automation Conference, pp. 284–290 (1997)
2. Kernighan, B.W., Lin, S.: An efficient heuristic procedure for partitioning graphs. Bell Syst. Tech. J. **49**, 291–307 (1970)
3. Jiang, Z.W., et al.: NTUPlace2: a hybrid placer using partitioning and analytical techniques. In: ISPD, pp. 215–217 (2006)
4. Subbaraj, P., SaravanaSankar, S., Anand, S.: Parallel genetic algorithm for VLSI standard cell placement. In: IEEE International Conference on Advances in Computing, Control, & Telecommunication Technologies, ACT 2009, pp. 80–84 (2009)
5. Myung-Chul, K.: SimPL: an effective placement algorithm. IEEE Trans. Comput. Aided Des. Integr. Circ. Syst. **31**(1), 50–60 (2012)
6. Yang, X.S.: Nature-Inspired Metaheuristic Algorithm. Luniver Press (2008)
7. Marichelvam, M.K., Prabaharan, T., Yang, X.S.: A discrete firefly algorithm for the multi-objective hybrid flowshop scheduling problems (2012)
8. Apostolopoulos, T., Vlachos, A.: Application of the firefly algorithm for solving the economic emissions load dispatch problem. Int. J. Combinatorics, Article ID 523806 (2011)
9. Kwiecien, J., Filipowicz, B.: Firefly algorithm in optimization of queueing systems. Bull. Pol. Acad. Sci. Tech. Sci. **60**(2), 363–368 (2012)
10. Khadwilard, A.: Application of firefly algorithm and its parameter setting for job shop scheduling. J. Ind. Technol. **8**(1) (2012)
11. Vijay, G., Marimuthu, S., Sait, A.N.: Comparison of firefly algorithm and artificial immune system algorithm for lot streaming in m-machine flow shop scheduling. Int. J. Intell. Syst. **5**, 1184–1199 (2012)
12. Sharma, P.K., Kaur, M.: A discrete firefly algorithm for VLSI circuit partitioning. In: 2014 International Conference on Electronics and Communication Systems (ICECS), pp. 1–4 (2014)

A Security Model for Protecting Virtualization in Cloud Computing

Saurabh Singh, Pradip Kumar Sharma, and Jong Hyuk Park(✉)

Department of Computer Science and Engineering,
Seoul National University of Science and Technology,
Seoul 139-743, Republic of Korea
{singhl989, pradip, jhparkl}@seoultech.ac.kr

Abstract. Virtualization is the key component of cloud computing that refers to the abstraction of sharing resources. The basic idea is to implement the virtualization in cloud computing environment because of its flexibility, scalability, its cost-reducing and resource utilization. Instead of many good attributes virtualization still facing with security flaws like availability, mutual authentication, potential attack in the virtual network, DoS attack on virtual servers and storage also. In the paper, we propose a general security model to protect the virtual environment and also discuss the flow of data which is monitored by hypervisors.

Keywords: Cloud computing · Virtualization · Attack

1 Introduction

Due to recent trends in IT like cloud computing and server consolidation, virtual techniques are gaining more and more importance. It has emerged out as a comer stone of the IT field. A cloud provider utilizes the virtualization techniques by combining the self-service abilities for sharing resources with the help of IaaS. Instead of a key component of cloud, virtualization faces many security problems, for example, there are many servers are running on the single physical machine and unfortunately, RAID controller take out all the hard disks causes data loss and leakage of information [1]. The problems and security issue are still present in different types of virtualization process [2].

So we need to improve the security concerns in different virtualization level, thus, we propose a model which deals with security problems and protect it in each phase of virtualization.

Although the wide application of virtualization, the shared use of devices like routers, servers, storages and many applications like OS and communication channels leads to a series of security concerns. Here we discuss the opinion of some author regarding security issue in different type's virtualization.

Network virtualization: Sriram Natrajan et al. explorer the security issues in network security. The paper explained while developing the infrastructure component which provides logical isolation between virtual networks and these systems often more

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_60

widely deployed which deals some security problems against malicious developer and users [2]. It basically discusses the potential attack and relationship between entities to illustrate the security issue while design and implementing the virtual networks.

OS virtualization: OS layer virtualization is also identified as a single kernel image which implements by running instances of same operating system parallel [3].

Server virtualization: A virtual server creates special security challenges like vulnerability in hypervisor, sometimes remote access problem and server reliability across virtual machines.

Storage virtualization: It is the form of resource virtualization which pools the physical storage from multiple network storage and appears to be a single storage device which managed from common central console. It increases storage utilization and it does not affecting application availability by adding or deleting storage data.

2 Proposed Virtualization Security Model

The most common and outstanding feature of cloud computing is virtualization. In a virtual environment so many virtual machines are running together and managed and monitored by hypervisor administrator. On each level of virtualization many security vulnerabilities are present [4], to remove these security problems we propose a new security model in a virtual environment as shown in Fig. 1 that represents the logical separation on each level of virtualization.

2.1 Analysis and Discussion

The security solution of virtualization is heavily dependent on individual security and safety of each component to guest OS's, applications, servers and storage from host OS and hypervisor [5]. The proposed model applied terminal authentication to authenticate the end user to enter into a virtual environment. The encryption methodology is applied in a virtual network to protect against an attacker, it is very important to secure virtual networks because of its widely and complicated structure.

Virtual server secures the malicious multitenant and unauthorized users, it resolves the issue by clearing virtual instance memory and prior loading new instance memory along with new CPU preventing cross reading of memory against malware instance. The basic approach to at OS level is to redirect the request from the virtual machine to local resource partition on a physical machine. This model ensures the security of communication between virtual OS and with client machine using cryptography. To protect the virtual storage encryption with the backup plan is available in proposed model.

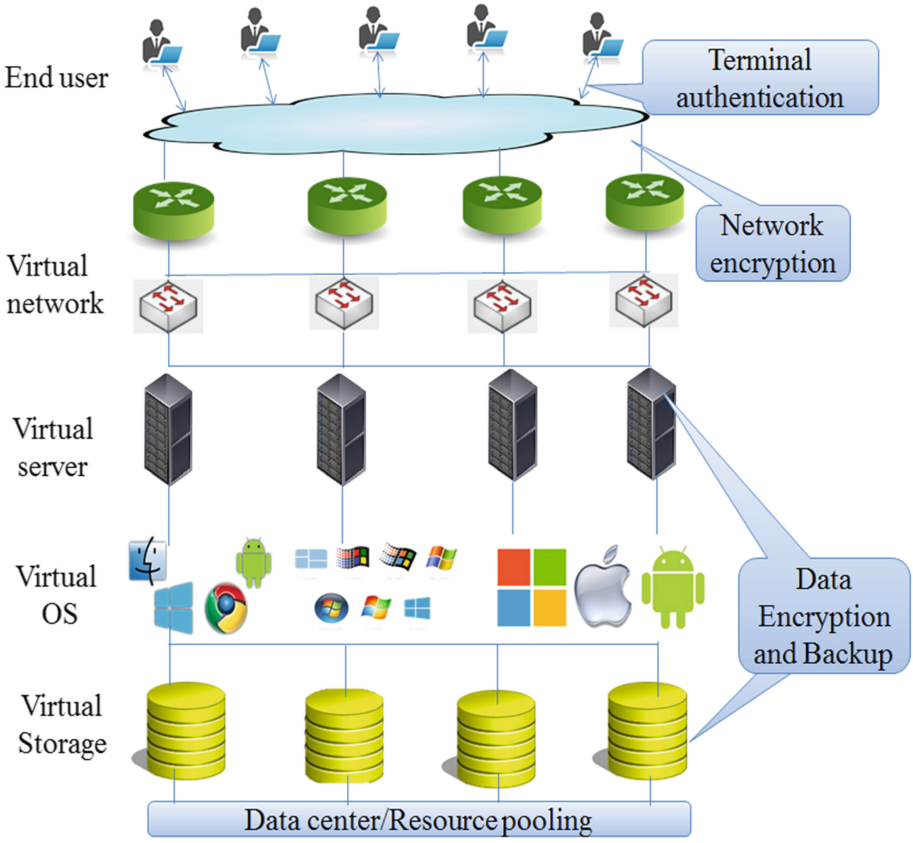


Fig. 1. General virtualization security model.

3 Conclusion

Virtualization in the cloud allows the workload to be deployed through provisioning of the virtual and physical machine. The paper has briefly presented the security flaws in different types of virtual machine environment. The security model helps to protect the virtual environment against malicious activity by full-filling the security parameters. Many virtual OS are running in parallel is protected by cryptographic techniques and avoid the problem of mutual authentication. The suggestive way for the issue and challenges in the enterprise is to maintain, updating patch of software and for accessing every time everywhere of decentralized application should be considered before transferring the data.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Anand, R., Saraswathi, S., Regan, R.: Security issues in virtualization environment. In: International Conference on Radar Communication and Computing (ICRCC), pp. 254–256, December 2012
2. Natarajan, S., Wolf, T.: Security issues in network virtualization for the future internet. In: International Conference on Computing, Networking and Communications (ICNC), pp. 537–543. IEEE, January 2012
3. Sgandurra, D., Lupu, E.: Evolution of attacks, threat models, and solutions for virtualized systems. J. ACM Comput. Surv. (CSUR) **48**(3), 46 (2016)
4. Kedia, P., Nagpal, R., Singh, T.P.: A survey on virtualization service providers, security issues, tools and future trends. Int. J. Comput. Appl. **69**(24), 36–42 (2013)
5. Loganayagi, B., Sujatha, S.: Enhanced cloud security by combining virtualization and policy monitoring techniques. Procedia Eng. **30**, 654–661 (2012)

Effective Pre-processing Methods with DTG Big Data by Using MapReduce Techniques

Wonhee Cho[1] and Eunmi Choi[1,2(✉)]

[1] Graduate School of Business IT, Kookmin University, Seoul, Republic of Korea
danylight@gmail.com, emchoi@kookmin.ac.kr
[2] Department of Financial Information Security, Kookmin University, Seoul, Republic of Korea

Abstract. A huge amount of sensing data is generated by a large number of pervasive IoT devices. In order to find a meaningful information from the big data, pre-processing is essential, in which many outlier data need to be removed because those are deteriorated as time passes. In this paper, big data pre-processing methods are investigated and proposed. To evaluate the pre-processing methods for accurate analysis, we use collection of digital tachograph (DTG) data. We obtained DTG sensing data of six-thousand driving vehicles over a year. We studied five kinds of pre-processing methods: filtering ranges, excluding meaningless values, comparing filters from variables, applying statistical techniques, and finding driving patterns. In addition, we developed MapReduce programming using a Hadoop ecosystem, and deployed a big data to perform pre-processing analysis. Out of the pre-processing steps, we confirmed the proportion of DTG sensing data including any errors is up to 27.09 %. In addition, we approved that outlier data can be well detected, which is difficult to detect through simple range error pre-processing.

Keywords: Big data · Pre-processing · Sensor · IoT · MapReduce

1 Introduction

Sensing devices are increasing in popularity, and a large amount of sensing data is generated by the pervasive IoT devices. Analysis using IoT data is also increasing by development of data mining and machine learning. Various kinds of outlier data are generated, as those devices usually have mobility and battery power shortage and deterioration due to exposure to a variety of environments according to IoT device features. Thus, filtering and correcting are necessary for such outlier data [4]. DTG has by law been installed on most commercial vehicles since 2011 to reduce traffic accidents in Korea [1]. A large amount of data including GPS trajectory is gathered from

E. Choi—This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the IT/SW Creative research program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2013-H0502-13-1071).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_61

DTG. DTG is an apparatus for recording the operating history of the vehicle, which is similar to a black box of an aircraft [2].

In this paper, big data pre-processing methods are investigated and proposed. To evaluate the pre-processing methods for accurate analysis, we use collection of digital tachograph (DTG) data. We obtained DTG sensing data of six-thousand driving vehicles over a year. We studied five kinds of pre-processing methods: filtering ranges, excluding meaningless values, comparing filters from variables, applying statistical techniques, and finding driving patterns. Pre-processing was also implemented and performed by building a Hadoop ecosystem environment and MapReduce features of distributed processing in Java [3].

2 Related Research Work

- **Pre-processing methodology**

DTG is a kind of IoT device for sensing data. Pre-processing is essential to this data, as noises are quite great and unstable in the process of gathering and transferring. Generally, it may cause problems such as degrading data quality. IoT data is also generated in large volumes by nature. Thus, pre-processing is an essential data mining techniques [7, 8]. To make a good estimation model requires good quality data. Performing a data status check to find the data in the problem early and determining appropriate data processing and cleanup steps are important for preventing the "Garbage in Garbage out" (i.e., poor data drives poor analysis) phenomenon and increasing the data quality to improve the ultimate performance model. Pre-processing can be divided into five steps: data cleaning, data integration, data transformation, data reduction, and data discretization. The detailed operation of each process is organized as follows: [5].

- *Data cleaning*
  - *Filling of missing values*
  - *Noisy data smoothing*
  - *Confirming or removing outliers*
- *Data integration*
  - *Data integration, converting consistent data type*
- *Data transformation*
  - *Normalization*
  - *Aggregation*
  - *Summarization*
- *Data reduction*
  - *Reducing size, but with identical analysis result*
- *Data discretization*
  - *A kind of data reduction*
  - *Numerical values converted into categorical values Conversion into categorical values*

- **Vehicle data pre-processing**

Silva introduced a pre-processing method using GPS trajectory data before estimating fuel consumption from his paper, which is similar to this study [6]. In his study, a filtering method is use to remove outliers, missing values, unrealistic data, low reliable data, and a very small amount of data. Removal criteria of his study were as follows [6].

(1) *points with missing values;*
(2) *points with GPS accuracy* < *50* m;
(3) *points with inclination out of the interval [−15; 15]*^{o};
(4) *points with acceleration out of the interval [−4; 4]* m/s^2;
(5) *points with fuel consumption* > 100 L/100 km;
(6) *points where the difference of speeds given by GPS and OBD is greater than* 3 m/s.

3 Pre-processing Design for Sensing Data

Because DTG data contains vehicle sensing information, it exhibits outliers, such as a large amount of errors in the generation and transmission processes. In this study, statistical values used in the filtering criteria are the driving records, which are instigated and ended by vehicle ignition in the journey unit. Filtering criteria were studied in five kinds of methods, as shown in Table 1.

Table 1. DTG data filtering methods

Method	List	Cause
1. Filtering range errors	Statistical number that exceeds the basic range of machinery	Sensor error
2. Excluding meaningless value	Meaningless value of statistics on vehicles in maintenance or idle within the basic range	Sensor error or idling, non-driving, short driving without statistical validity
3. Comparing filters from variables	Filtering by comparing the values between the two variables	Sensor error or idling, short-distance driving without statistical validity
4. Applying statistical techniques	Pearson Correlation, error detection by using standard deviation, etc.	Maintenance, sensor error, idling, short driving
5. Finding driving patterns	Error detection using a driver's driving pattern, like rapid operating	Sensor error

- **Range Error (Basic Range Value Criteria Filtering)**

Sensing information includes a basic Range. Its value may be absent or meaningless, depending on the nature of the equipment. In this study, outliers are detected

and filtered by range values, which are realistically difficult to generate. The reference values are as follows.

① *Speed: > 163*
② *PM: > 4000*
③ *Brake signal: confirming the value besides 0 and 1, or one of the incoming brake signals 0 or 1.*
④ *GPS X,Y: confirming territory of the Republic of Korea (125.8 < x < 131.0, 33.0 < y < 39.0)*
⑤ *Dir(direction): 0–360*
⑥ *Acc X,Y (acceleration): −100 ∼ +100*

- **Meaningless values**

Although the data is in a limited range, meaningless values, such as data from maintenance and idle vehicles are excluded. For example if the total speed is zero, it is estimated to be sensor failure or stationary vehicle. If the RPM is zero, it is a case of the key being switched on but the engine not having started status. In this case, the value is recorded, but it is meaningless record data. Also, cases in which total driving time is less than 60 s should be excluded, because the data is also subject to the distortion of statistics.

- **Comparison filtering from variables**

Though each variable comes within a basic range, errors can be included due to mechanical failure or communication error. These errors can be detected by comparison of the variables that are associated with each other.

① *Compare car speed signal and GPS speed conversion value (Speed sensor or GPS device error)*
② *In cases where acceleration is less than 5 and maximum speed greater than 100* km/h *(Speed sensor error)*
③ *In cases where average speed is greater than 5 and less than 200* m *of GPS conversion distance (Idle or short distance drive)*

Speed value of the vehicle speed sensor and converted velocity value of GPS moving distance should be similar under normal conditions. However, if there is a lot of difference, it could be due to speed sensor or GPS sensor error. In addition, if acceleration is less than 5 and maximum speed is greater than 100 km/h, speed sensor error could be estimated. If average speed is more than 5 and GPS conversion distance is less than 200 m, meaningless data for analysis could be estimated as idling or short-distance driving.

- **Filtering using Statistical techniques**

Some errors are difficult to detect, even if comparison between variables is possible. We studied this kind of error detecting method using some statistical techniques

① *Pearson correlation between speed and RPM (in cases less than − 0.7)*
② *Dangerous driving pattern detection analysis*
③ *In cases where speed/RPM increase/decrease standard deviation is larger than or equal to zero.*

Speed and RPM values generally have a correlation greater than 0.5. However, some data may have strong inverse correlations, such as less than −0.8 or −0.9. Looking at these data, there are some data, such as machine abnormality or use as a maintenance vehicle that cannot use the statistical information. Accordingly, it can be filtered using the correlation value. In addition, there are some repeat patterns that represent rapid acceleration and deceleration within a one-second gap, even though these data are in the basic range. In this case it also turns out to be a device error. If the speed and RPM increase or decrease standard deviations are zero, speed and RPM sensor error is found. These data may also be filtered by the detection using the standard deviation value.

4 Big Data Analysis Design and Development

- **MapReduce for pre-processing**

A single server has limitations in pre-processing for the large scale sensing data. Thus, in this study, pre-processing was designed and implemented using Hadoop MapReduce. Key Values are vehicle number and date in the Map process. Value Range is the Count in excess of each variable as in Table 2.

Table 2. Key/value of the Map process

Key (CarNum + Date)	Value (OPCode + Count)
BS99F11752 + 140425	Speed + 1
BS99F11752 + 140425	GPSX + 1
:	:
CN99G01208 + 140429	RPM + 1

Table 3. Key/value of the Reduce process

Key (CarNum + Date)	Value (SpeedCnt, RPMCnt, ….., BrakeSign)
BS99F11752 + 140425	0, 30, 0, 0, 0, 26, 0, 0, 0, 0, 0, 0, 2, 118, 1920, 0, 0, 0
BS99F11752 + 140425	1, 35, 0, 0, 0, 31, 0, 0, 0, 0, 0, 0, 2, 115, 1867, 0, 0, 0
:	:
CN99G01208 + 140429	0, 77, 42, 12, 0, 23, 0, 0, 0, 0, 0, 0, 2, 0, 0, 0, 0, 0

Statistical results are calculated by summing the Value of the Key value in the Reduce process as in Table 3.

5 Experimental Result

Table 4 indicates pre-processing results analyzing DTG total driving data of one year in the Journey unit

The number of variables in the GPS error, which is greater than the basic Range. This corresponds to 3.54 % of the total. Brake error signal corresponds to 10.05 % of the total, which is the case with only one of the incoming brake signals, 0 or 1. It covers

Table 4. Pre-processing results

Type	Error name	Outlier criteria	ErrRate (%)	Estimated reasons
1. Range error	GPS error	GPS signal include 0 or out of Korea territory	3.54	GPS sensor error
	Brake signal error	Avg brake sig. is 0 or 1	10.05	Brake sensor error
	Speed error	highSpeed > 163	4.39	Speed sensor error
	RPM error	highRPM > 4000	3.37	RPM Sensor error
	Acceleration error	Average Accs is 0	5.54	Acc Sensor error
2. Meaningless values	Total Speed is zero	highSpeed = 0	7.08	Sensor error or idling
	Total RPM is zero	highRPM = 0	2.83	Sensor error or not start-up
	Driving time is less than 60 s	secCnt < 60	0.78	Very short driving
	Excessive idling	idleCnt > 1000	0.08	Speed Sensor error or excessive idling
3. Comparison filtering from variables	Compare car speed and GPS speed	avgDGpsSpeed<20 or >50	4.05	GPS Sensor error Speed Sensor error
	Acc less than 5 and maximum speed greater than 100 km/h	AccX < 5 and highSpeed > 100	0.24	Sensor error
	Average speed greater than 5 and GPS distance less than 200 m	avgSpeed > 5 and sumGpsDist < 200	1.80	Idling or very short driving
4. Filtering using Statistical techniques	Correlation between speed and RPM is less than −0.7	Correlation < −0.7	0.04	Vehicle maintenance or sensor error
	RPM Standard Deviation	SD_Rpm > 4000 and SD_Rpm < 100	6.49	Sensor error or excessive idling
	Speed increase standard deviation	SD_PSpeed > 10	1.57	Sensor error
	Speed decrease standard deviation	SD_DSpeed > 10	0.54	Sensor error
5. Driving Pattern	driver's driving pattern	hSUpCnt > 40	0.00	Speed Sensor error
Sum (Redundant error calculate one)			27.09	

the most cases. The results indicate speed sensor error 4.39 %, RPM sensor error 3.37 %, acceleration sensor error 5.54 %, and overall sensor error DTG at least 10 %. Value compared to vehicles with GPS Speed is 4.05 %, in which case it is GPS or Speed sensor error. The total number of errors is 266,675, that is, 27.09 % of the 1,042,027-journey data. This indicates the possibility of including errors in the multiple paths such as sensing, communication pathway, and server storage. Our pre-processing methods achieved the highly qualified error-filtering.

6 Conclusion

In this paper, the pre-processing methodology was investigated, and big data pre-processing was also studied by building MapReduce. Handling data of IoT devices, pre-process is essential because large amounts of outlier data are generated. In this study, the proposed methodology was analyzed using 1-s unit DTG data of 6000 Korean driving vehicles over a year. We studied five kinds of pre-processing methods (range error, meaningless value, comparison filtering from variables, statistical techniques, and driving pattern), and applied such methods by developing MapReduce codes using Hadoop ecosystem.

We confirmed that the proportion of driving data including one error is up to 27.09 %. We also confirmed that outlier data can be detected, which is difficult to detect through simple Range Error check. Improvement of IoT sensing pre-processing methodology was confirmed using statistical techniques. Further research is required to improve sensing data processing in various other fields.

References

1. Lee, S.J., Lee, C.: Short-term impact analysis of dtg installation for commercial vehicles. J. Korea Inst. Intell. Transp. Syst. **11**(6), 49–59 (2012)
2. Standard Specification of DTG (Ministry of Land, Infrastructure and Transport, KS R 5072, February 2009. (in Korea)
3. White, T.: Hadoop: The Definitive Guide. O'Reilly Media, Inc., Sebastopol (2012)
4. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**, 2787–2805 (2010)
5. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Elsevier, New York (2011)
6. Vilaça, A., Aguiar, A., Soares, C.: Estimating fuel consumption from GPS data. In: Paredes, R., Cardoso, Jaime, S., Pardo, Xosé, M. (eds.) IbPRIA 2015. LNCS, vol. 9117, pp. 672–682. Springer, Heidelberg (2015). doi:10.1007/978-3-319-19390-8_75
7. Cho, W., Choi, E.: A GPS trajectory map-matching mechanism with DTG big data on the HBase system. In: The 2015 International Conference on Big Data Applications and Services, October 2015
8. Cho, W., Choi, E.: Rural traffic map coverage extension using DTG big data processing. J. Inf. Technol. Architect. **12**, 51–57 (2015)

Security Requirements and Countermeasures for Secure Home Network in Internet of Things

Seo Yeon Moon, Saurabh Singh, and Jong Hyuk Park(✉)

Department of Computer Science and Engineering, Seoul National University of Science and Technology, Seoul 139-743, Republic of Korea
{moon.sy062l, singhl989, jhparkl}@seoultech.ac.kr

Abstract. The home network share the data with each other to configure a network after connecting many devices that wired or wireless in the home. Recently the home network has increased as internet of things service extension. But this is not achieved strong security because due to this, each device has the function deterioration problem in home network. Therefore it is necessary to determine the wired or wireless network security vulnerabilities and countermeasures to prevention for cybercrime privacy leakage. In this paper analyzes the factors of home network vulnerabilities and middleware security in internet of things. Also it presents a security vulnerability improvement.

Keywords: Home network · Security vulnerability · Middleware · Internet of things

1 Introduction

With the development of information communication technology, humans are increasing interest in more diverse and convenient service. As common networks and devices, we can easily configure the smart home environments in home through the development of the internet of things (IoT). In general, home network connects various devices as network that wired or wireless in the home. And it constitutes a home network environment for share the communication such as data, printer, and internet. Home network in internet of things is capable of communication between the devices.

We apply security solutions to network devices in the home, it appears the impact of deterioration is very large. So it is difficult to enforce strong security and is exposed to cyber-crime, such as information disclosure, alteration by using vulnerability. Especially when exposed to security threats, the risk is very large. Because in all device in internet of things is configured a network and it has the data flow. We need to know about security vulnerabilities and countermeasures that wired and wireless to prevent privacy leakage and cyber-crime in home network. This research analyzes the factors of home network vulnerabilities and middleware security in internet of things to improve the security network and applications from smart home network. And this suggests a vulnerability improvement.

Internet of things are defined “This is a technology anytime and where the network connection between people and things.” in the ITU [1]. This is a new paradigm that described various techniques have been integrated such as sensor recognition

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_62

technology, network communication technology, embedded technology, web and application services technology etc. [2].

Internet of things is similar, but different ubiquitous and machine to machine (M2M). Because was fused to the existing communication network system. And it was systematized and extended exceed all device capable of M2M. In particular, unlike M2M communication, this is generally applied to the web in a network environment established between things. And it is actively utilizing big data, data mining, etc. Also it has the potential to generate other content. Figure 1 shows the approaches on internet of things solutions. In summary, internet if things can be defined as a new concept that combines the complex technology based on communication function in variety devices.

Internet of things already been studied for this component in a number of ways [3]. There are intelligent, structure, complex system, size, time, space. This is a very important indicator g for each step considering such as design, development, and evaluation the actual implementation. When utilizing these elements, we can easily and efficiently build Internet of things system. And it can be effectively used in maintenance and repair.

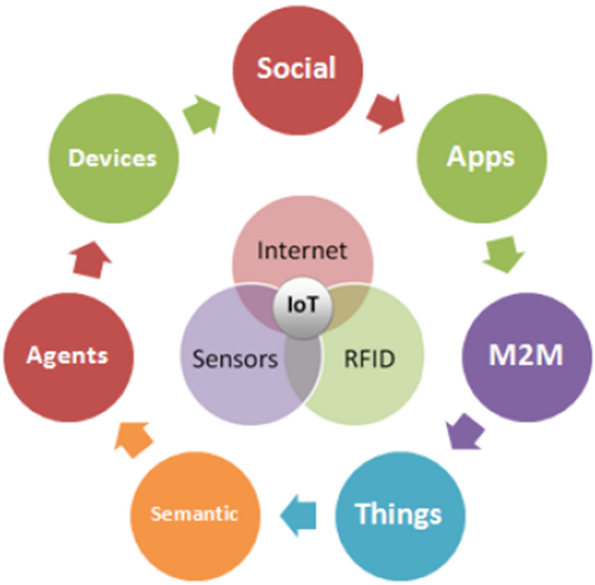


Fig. 1. Approaches on internet of things

2 Home Network

The home network is to establish a mutual network utilizing a variety of information appliances in the home. In the internet of things environment, Smart devices in home network has the mutual communication using network that wired or wireless. This means that the build environment that can be interconnected through the Internet from outside. This configuration of a home network is shown in Fig. 2.

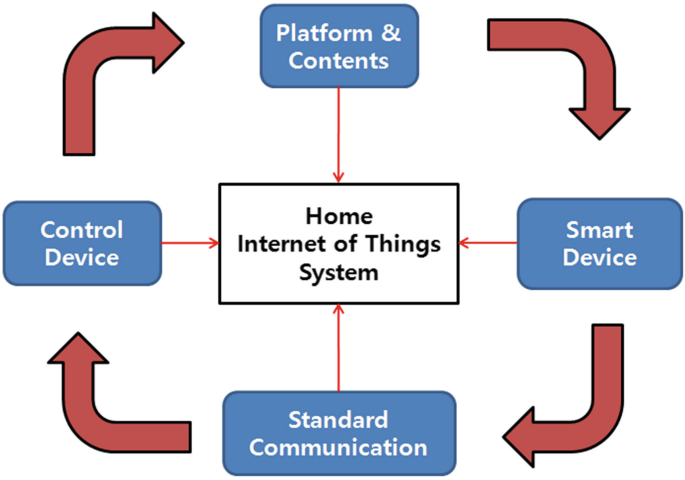


Fig. 2. Home internet of things system component

2.1 Home Network Middleware

In this section, we analyzed the characteristics of the middleware. The middleware provides basic security functions for control between devices of the home network. Also this has been made standard about related security. In particular, the function of middleware are more various.

Middleware is generally located between the OS (Operating System) and application environment. This is also referred to as a software layer that provides the solution of situations that can be generated from connectivity issues, compatibility, security, reliability between heterogeneous devices [4].

Middleware is in charge of various functions. Such as internet of Things environmental services support common interfaces, network connectivity between heterogeneous devices, each sensor data acquisition and processing, device control and data conversion, and dynamic resource management, etc. Among the core of middleware is which will connect the network between different devices. It allows to convert the data to enable communication between different protocols.

Generally Middleware is software that mounted to the home server for a home network. This allows an intermediary for all devices connected able to exchange control information and various kinds of multimedia information to each other in home network. One of t network middleware security features in home network is usually the authentication and access control of the devices. In addition, this provides user authentication, message integrity and confidentiality. Usually it does not have strong security due to the fear of poor performance.

2.2 Home Network Security Vulnerabilities

It is possible to use various security techniques. However, due to problems with the computing power, it does not have the strong security features that can solve security vulnerabilities. Home network vulnerability in the wired and wireless are shown in the Table 1 below.

Table 1. Home network vulnerability

Network	Technology	Security vulnerability
Wired	IEEE 1394	– Outlet and the modulation caused by data transmission and reception
	PLC	– Modulation and outlet that control information/data transmission
	HomePNA	– Built-in SNMP – Authentication and privacy service vulnerability – A Dos and traffic analysis vulnerability
	USB	– No retransmission – Limited low-speed transfer mode (keyboard and mouse)
Wireless	WLAN	– Plaintext exposure – Potential threat Ddos
	UWB	– Threat based on wireless openness
	Wireless LAN	– Need the high transfer rates supported
	Wireless 1394	– Potential threats undercover
	Zigbee	– Blue sniffing – Blue bugging – Blue jacking

3 Conclusion

As Internet of things penetration, home network environment configuration was inevitable. However, the attack technology is increased by the home network with the development of technology. As a result, an attacker appear in a variety of attack types. Therefore, we analyze the security requirements for the improvement of security vulnerabilities and problems in the home network. And we show a countermeasure for it.

The future Internet of Things and home network will be increased. At the same time, has the function of middleware is extended gradually increasing importance. But many researchers do not strongly discuss the security countermeasures of the middleware. Therefore, we should discuss a security vulnerability by providing a common security standards to be satisfied in the future. And it must respond to possible cyber-crimes and privacy leakages that are occur at any time. Through threats and requirements for research in this paper, it should develop specific and systematic research and security solutions. And building a more secure home network will have to be made.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP (Institute for Information & communications Technology Promotion)

References

1. International Telecommunication Union (ITU): Internet of Things, International Telecommunication Union Internet Reports (2005)
2. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2575–2806 (2014)
3. Sundmaeker, H., Guillemin, P., Friess, P., Woelffle, S.: Vision and challenges for realizing the internet of things. European Commission Information Society and Media (2010)
4. Issarny, V., Caporuscio, M., Georgantas, N.: A perspective on the future of middleware-based software engineering, pp. 244–258. IEEE Computer Society (2007)

Machine Learning for Trajectory Generation of Multiple-pedestrians

Hye-Yeon Yu(✉), Young-Nam Kim, and Moon-Hyun Kim

College of Information and Communication Engineering,
Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu, Suwon-si
Gyeonggi-do 440-746, Republic of Korea
{yu0529,hwarangjin,mhkim}@skku.edu

Abstract. In this paper, we provide an algorithm for generating a trajectory in real time to identify the pedestrians. Typically, the contours for the extraction of pedestrians from the foreground of images are not clear due to factors including brightness and shade; furthermore, pedestrians move in different directions and interact with each other. These issues mean that the identification of pedestrians and the generation of trajectories are somewhat difficult. We propose a new method for trajectory generation regarding multiple pedestrians. The first stage of the method distinguishes between those pedestrian blob situations that need to be merged and those that require splitting, followed by the use of trained decision trees to separate the pedestrians. The second stage generates the trajectories of each pedestrian by using the point-correspondence method; however, we introduce a new point correspondence algorithm for which the A* search method has been modified. By using fuzzy membership functions, a heuristic evaluation of the correspondence between the blobs was also conducted. The proposed method was implemented and tested with the PETS 2009 dataset to show an effective multiple-pedestrian-tracking capability in a pedestrian interaction environment.

Keywords: Artificial intelligence · Multiple-target tracking · Machine learning · Decision tree

1 Introduction

For a long period of time, tracking has been an important theme in terms of computer vision, and it remains a challenging task. Computer-vision tracking involves the detection of the moving regions of images, the separation of foreground pedestrians from the background, and movement prediction. The tracking of a pedestrian is straightforward, as the pedestrian can easily be found in each frame, and the elapsed time can be used to determine the trajectory of the detected pedestrian; however, the generation of the trajectories of multiple tracked pedestrians is more complicated. During the tracking of multiple pedestrians, it is possible to mistakenly identify one of the other pedestrians as the target pedestrian, or to miss the detection of a pedestrian due to inter-pedestrian interactions such as occlusions.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_63

The objects tracking and detection is the most important and essential step in computer vision. The main target of object tracking is to detect and track the moving objects such as pedestrians.

Occlusions and overlaps occur very commonly in pedestrian scenes. Depending on incomplete object detection, tracking algorithm is more complex. Calculation for connecting the tracklets belonging to the same individual is NP-complete problem.

Yilmaz et al. [1] categorize the tracking methods on the basis of the object and motion representations used. There are two main processes in motion-based recognition: One is a motion representation, another is the extraction of motion information from a sequence of images. In general, for extracting motion information, using the two methods as follows: motion correspondence and optical flow [2]. Eom et al. [3] proposed a heuristic search-based motion correspondence algorithm using fuzzy clustering. They focused on the motion corresponding algorithm. And in the experiment, corresponding points is a preprocessing. In this paper is an improved version of their pre-processing phase. Pasquale Foggia et al. [6] proposed system is able to simultaneously track single objects and groups of objects. The object classifier module classifies each single blob into different classes.

2 Proposed Method

The diagram of the proposed method is shown in Fig. 1. We separated the foreground from the background by using a Gaussian mixture model whereby the separated foreground consists of blobs. Each pedestrian blob is represented by features such as area, perimeter, aspect ratio, center of mass, average color, and the distance to the closest blob in the previous frame; based on training-set learning, these features are used for the construction of the decision trees. The resultant learned decision trees are used to classify the blobs as either "split" or "merged," and the splitting and merging processes assign blobs to each pedestrian.

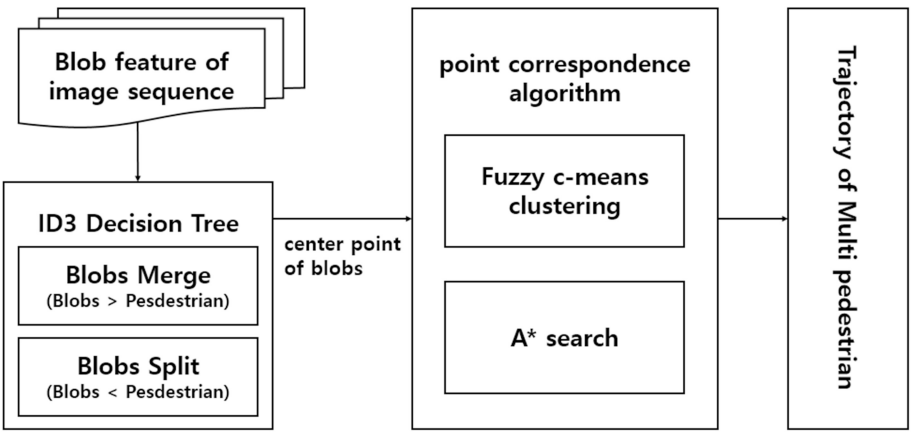


Fig. 1. Diagram of the proposed method

To associate a blob in the current frame with a tracked pedestrian from a previous frame, we devised a modified A* heuristic search algorithm. For each blob, the search for the most-probable pedestrian from the previous frame is based on the Euclidean distance and the difference between the velocity angles of the blob and the pedestrian. The required heuristic function that computes the possibility that a tracked pedestrian corresponds with a blob was designed using a fuzzy-C-means (FCM) clustering algorithm. It is the association of the blobs with the pedestrians over the entire frame sequence that produces the trajectories, each of which is a trajectory for a single pedestrian.

2.1 ID3 Decision Trees

The ID3 decision tree is an information-theoretical algorithm invented by Ross Quinlan [7]. Each iteration of the ID3 involves the attainment of the largest information gain (or smallest entropy value), and a probability of p is determined from the occurrence frequency. The entropy of a given B_M is computed as follows:

$$E(B_M) = -\sum_{C_i \in C} p(C_i) \log_2 p(C_i) \tag{1}$$

where C is a set of classes in B_M that is either of a positive (merging) class or a negative (not-merging) class, and p(C_i) is the proportion of the number of pairs labeled as class C_i in B_M to $|B_M|$. The *information gain(IG)* from the partitioning of B_M according to the value of the feature f_M is computed using Eq. (2), as follows:

$$IG(f_{M_i}, B_M) = E(B_M) - \sum_{v_j \in v_{M_i}} p(v_j) E(B_{v_j}^M) \tag{2}$$

In Eq. (2), $E(B_{v_j}^M)$ is the entropy of the subset $B_{v_j}^M$ that is constructed through the collection of pairs, each of which has a value of v_j for the feature f_{M_i}, from B_M. V_{M_i} denotes a set of values for f_{M_i}; therefore, $B_M = \bigcup_{v_j \in M_i} B_{v_j}^M$. The probability $p(v_j)$ is computed using Eq. (3), as follows:

$$p(v_j) = \frac{\left|B_{v_j}^M\right|}{\left|B^M\right|} \tag{3}$$

At each iteration of the construction of a decision tree, the feature that provides the maximum information gain is selected as a decision node. To deal with the numeric features, the C4.5 [8] ID3 algorithm is used.

2.2 Trajectory Generation

A* algorithm is a typical heuristic shortest path search algorithm. A* algorithm select the node with the minimum value of the values that have been calculated by guidance

function that uses a heuristic estimate to a depth first search [5]. The evaluation function in search node n is given in Eq. (4).

$$f(n) = g(n) + h(n) \tag{4}$$

With g(n): path cost to the n node from the starting node. h(n): estimated cost of the path to the target node from the n node. h(n) usually use the manhattan distance. The Algorithm progress that the node with a value of shortest guidance function in node to be connected form the starting node is to be a node that make up the shortest path, but it will end by selecting the target node. A* algorithm has the advantage that a search time is shortened in accordance with finding the target node without comparison with other paths. However, in a complex and unstable network, there is the potential for that to fail the search for shortest path. The number of clusters and the initial cluster center value at the approximate point are calculated by using a Mountain method [5]. With many changes by the FCM algorithm repeatedly, the positions of the cluster centers that are determined roughly is determined the final cluster center. Using the degree of affiliation calculated by the FCM algorithm, cost of all correspondence are calculated then the cost is used in making the correspondence matrix. Finally, an optimal trajectory is generated by the correspondence matrix and the A* algorithm.

2.3 Experiment Results

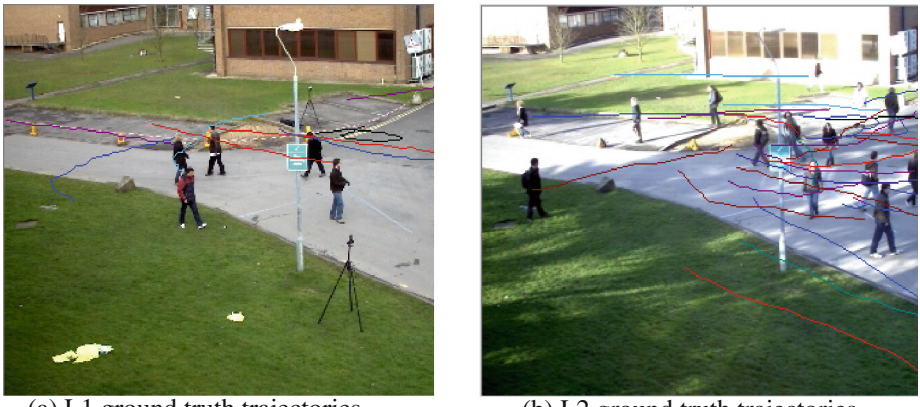
We used the Caviar "Crowd of four people meet, walk and split" sequence, PETS 2009 dataset S2 from the L1 sequence view 001 and the L2 sequence view 001 as shown in Fig. 2. The Caviar sequence is used only for tracking in clean environment. WEKA (Waikato Environment for Knowledge Analysis) version 3.7, written in JAVA, is an experimental tool used for machine learning [11] that we used for ID3 machine learning and the visualization of the decision trees. Blob detection and a trajectory-generation algorithm were implemented in OpenCV 2.4.2 with Visual Studio 2010.

(a) In the merging situation (b) In the splitting situation

Fig. 2. Merge or split situation

Figure 2 shows splitting or merging situations for the detected blobs in the image. Two pedestrians are detected as one blob, so the blob number 0 of Fig. 2 (a) needs to be split; furthermore, one pedestrian is detected as two blobs, so the blob number 0 and the blob number 1 of Fig. 2 (b) need to be merged.

Figure 3 shows the generated trajectories for the S2 views of the L1 and L2 sequences from the PETS 2009 dataset. The pedestrian group consists of nine people in the (a) L1 sequence and 33 people in the (b) L2 sequence. The pedestrians of (a) frequently change their moving direction in comparison with those of (b). The L1 and L2 sequences generated trajectories of 30 frames and 26 frames, respectively.



(a) L1 ground truth trajectories (b) L2 ground truth trajectories

Fig. 3. Multiple trajectories of pedestrians

3 Conclusion

Generating a trajectory in the multiple pedestrian have a lot of difficult factors. In the previous research, presented excellent results from trajectory generation of multiple target. The biggest problem for the application of previous research to multiple pedestrian tracking is the interaction between each other. We apply the ID3 algorithm to solve the problem of the interaction of the pedestrian. ID3 algorithm, the blobs overlap between the pedestrian was used to define the split and merge situations. Redefine the blobs to display a multiple-pedestrian in the experiment, to calculate the corresponding cost. Multiple-pedestrian trajectory is connection of the minimum cost of the node in A* search algorithm. Our experiments solved several constraints problems for the applied of point correspondence method to trajectory generation of multiple-pedestrian. The experimental results are classified as unclear blob splitting and merging situations. Thereafter, to generate a trajectory of each multiple-pedestrian as FCM based point correspondence method. The results of the research was successful in generating the trajectory of the multiple-pedestrian to the frequent interaction. In the future, to improve the accuracy of classification, via the ID3 algorithm for improved accuracy, generating the trajectory of each multiple-pedestrian with Datasets in diversity of experiments is expected to proceed to the next research challenge.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIP) (NRF-2014R1A2A1A11053902).

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: a survey. ACM Comput. Surv. (CSUR) **38** (4), 13 (2006)
2. Cedras, C., Shah, M.: Motion-based recognition a survey. Image Vis. Comput. **13**(2), 129–155 (1995)
3. Eom, K.-Y., Jung, J.-Y., Kim, M.-H.: A heuristic search-based motion correspondence algorithm using fuzzy clustering. Int. J. Control Automat. Syst. **10**(3), 594–602 (2012)
4. Yu, H.-Y., Kim, Y.-N., Kim, M.-H.: Temporal search algorithm for multiple-pedestrian tracking. KSII Trans. Internet Inf. Syst. **10**(5), 2310–2325 (2016)
5. Yager, R., Filev, D.: Approximate clustering via the mountain method. IEEE Trans. Syst. Man Cybern. **24**(8), 1279–1284 (1994)
6. Foggia, P., et al.: Real-time tracking of single people and groups simultaneously by contextual graph-based reasoning dealing complex occlusions. In: 2013 IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS), IEEE (2013)
7. Quinlan, J.R.: Induction of decision trees. Mach. Learn. **1**(1), 81–106 (1986)
8. Salzberg, S.L.: C4.5: programs for machine learning by J. Ross Quinlan. Morgan Kaufmann Publishers Inc, 1993. Mach. Learn. **16**(3), 235–240 (1994)
9. Berclaz, J., et al.: Multiple object tracking using K-shortest paths optimization. IEEE Trans. Pattern Anal. Mach. Intell. **33**(9), 1806–1819 (2011)
10. Hart, P.E., Nilsson, N.J., Raphael, B.: A formal basis for the heuristic determination of minimum cost paths. IEEE Trans. Syst. Sci. Cybern. **4**(2), 100–107 (1968)
11. Weka 3 - Data Mining With Open Source Machine Learning Software In Java. www.cs.waikato.ac.nz
12. Ferryman, J., Shahrokni, A.: PETS 2009 (2009). http://www.cvg.rdg.ac.uk/PETS2009
13. Homepages.inf.ed.ac.uk, CAVIAR Test Case Scenarios (2015). http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

A Deep Learning-Based Gait Posture Recognition from Depth Information for Smart Home Applications

Md. Zia Uddin(✉) and Mi Ryang Kim

Department of Computer Education, Sungkyunkwan University,
Seoul, Republic of Korea
ziauddin@skku.edu, mrkim@skku.ac.kr

Abstract. Gait posture recognition at smart environment is considered as a primary function of the smart healthcare nowadays. The significance of gait posture recognition is great especially for the elderly as it is one of the basic activities to promote and preserve their health. In this work, a novel method is proposed for gait posture recognition that utilizes Local Directional Patterns (LDP) for the local feature extraction of depth silhouettes. Then, DBN is trained to be applied for the posture recognition later. The proposed approach shows superior recognition performance over other traditional methods of gait posture recognition.

Keywords: Gait · LDP · DBN

1 Introduction

Recently, smart home is attracting a lot of attentions by many researchers to promote peoples' health [1–5]. It can play a key role to support the elderly or disabled people to support their life and looking after their health requirements. Among different human activity recognition at smart homes, gait posture recognition is considered to be a primary function nowadays [6]. In this regard, video cameras are being commonly used for various surveillance applications such as human computer interaction.

To represent the human body silhouette in images, binary silhouettes are the most commonly employed from which useful features are derived [7–12]. Agarwal and Triggs adopted binary silhouettes for human body pose recovery in [8]. Howe applied a binary silhouette-based approach to analyze human motion features in [9]. However, binary shapes are not efficient enough to describe human silhouette properly because of its limited flat binary pixel intensity distribution though they have been very much utilized. Besides, it seems extremely tough to represent the differences between the far and near parts of human body from binary shapes. To overcome such limitations of the binary shapes, depth shapes were proposed [13]. Depth shapes was successfully used in gesture recognition works such as [13]. In depth-based shape representation, the pixel values are assigned based on the distance to the camera. Hence, they can be more useful than binary shapes by providing better gait information in the images. Therefore, it would definitely be very wise to adopt depth shapes in gait posture recognition,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_64

which should allow us to come up with more robust gait recognition than binary shapes.

Regarding body shape features in gait, the very much common feature extraction method utilized so far is Principal Component Analysis (PCA) [10, 11, 13]. PCA extracts global features of human body in gait silhouettes and hence local feature extraction can provide more details regarding human body description in gait activity. In this regard, more robust human activity features can be obtained applying local feature extraction approaches such as Independent Component Analysis (ICA) and Local Binary Pattern (LBP). ICA is a higher order statistical approach than PCA and it produces local features that basically results in an improved recognition performance over PCA for human activity recognition [7].

LBP features basically focus on more detail than ICA as they have tolerance against illumination variances on the pixels in the images [14–17]. Hence, LBP are very robust features due to their tolerance against illumination variation. Besides, LBP features are very easy to compute. LBP was improved by Local Directional Pattern (LDP) later on [18]. Similar to LBP, LDP features also have the tolerance against illumination variations but LDP features represent much robust features as they include the direction information for each pixel [18]. Thus, LDP features should be adopted to represent human gait silhouette to apply for efficient gait recognition.

Deep Neural Network (DNN) is gaining a lot of attentions by many researchers these days as it can generate features from raw input [19]. Besides, DNN can overcome some limitations of basic perceptron where it is not able to perform in general pattern recognition. However, one limitation in DNN is it requires very much training time [19]. An improved version of DNN was proposed in 2004 which was called Deep Belief Network (DBN) [20]. DBN utilizes Restricted Boltzmann Machine (RBM) where the networks can obtain optimal solutions with much less training time than DNN. Thus, DBN is adopted in our work with LDP features to develop a robust gait posture recognition system.

2 Proposed Method

Robust features were obtained first using LDP to train a DBN, which is applied later on for recognition in testing. Figure 1 shows overall gait posture recognition architecture including training and testing via DBNs. A commercial depth camera was deployed to record depth images of different gaits where in the depth images, each pixel is represented as a gray value to represent shorter ranged pixels with brighter and longer ones with darker values. Figures 2(a) and (b) depicts some generalized depth shapes from a normal and abnormal gaits.

2.1 Local Directional Pattern (LDP)

Generally, Local Directional Pattern (LDP) uses an eight-bit binary code to each pixel in a depth image. The pattern is calculated by comparing the edge responses of a pixel from eight different directions. Kirsch edge detector is used in this work as the main

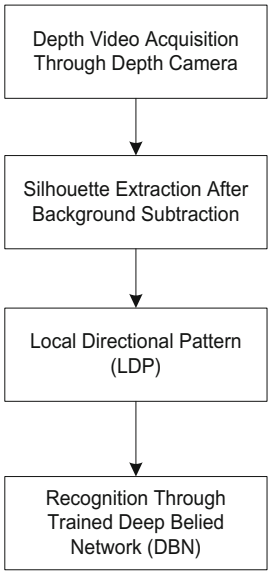


Fig. 1. Processes involved in the proposed gait posture recognition system.

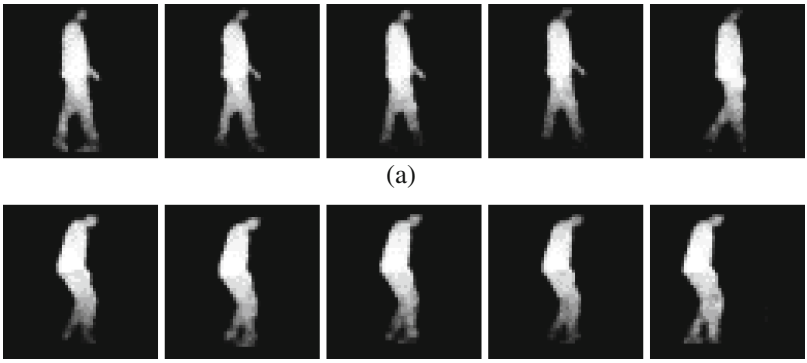


Fig. 2. A series of depth silhouettes of (a) normal and (b) abnormal gaits.

significance of this detector is it detects the edges more accurately than the others by considering eight neighbours. Thus, for a pixel in the image, eight directional edge response values $\{m_k\}$, $k = 0, 1, \ldots, 7$ can be computed by Kirsch masks. Figure 3 shows these Kirsch masks.

It would be interesting to know the p most prominent directions to generate LDP pattern for a pixel as all the directions cannot represent equal strength and hence low response directions can be ignored in this regard. The top-p directional bit responses b_k are set to 1 whereas the remaining bits are 0. At last, LDP code can be derived by (1).

$$\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}$$

East Direction M_0 Northeast Direction M_1 North Direction M_2 Northwest Direction M_3

$$\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}$$

West Direction M_4 Southwest Direction M_5 South Direction M_6 Southeast Direction M_7

Fig. 3. Kirsch edge masks in eight directions.

Figure 4 depicts an LDP code by considering the strength of top five directions. In our work, we adopted five top directions as other numbers more than that could not bring us better recognition performance in terms of recognition rates.

$$LDP_p = \sum_{k=0}^{7} D_k(m_k - m_p) \times 2^k, \quad D_k(a) = \begin{cases} 1 & a \geq 0 \\ 0 & a < 0 \end{cases} \tag{1}$$

where, m_p is the p-th most significant directional response.

Thus, an image was transformed to the LDP map using LDP code. The image textual feature is presented by the histogram of the LDP map where a bin can be defined as

$$T = \sum_{x,y} G\{LDP(x,y) = t\}, t = 0, 1, \ldots n-1 \tag{2}$$

where n is the number of the LDP histogram bins (normally $n = 256$) for an image G. Then, the histogram of the LDP map is presented as

$$H = (T_0, T_1, \ldots, T_{n-1}). \tag{3}$$

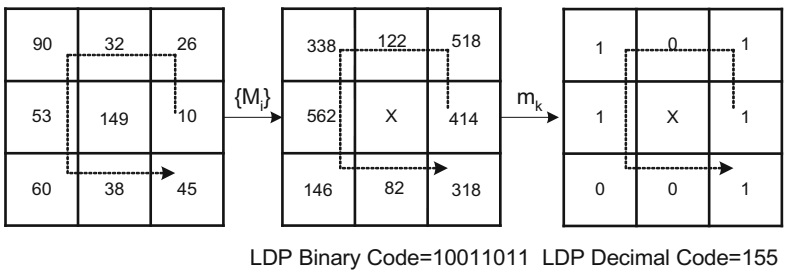


Fig. 4. LDP code.

2.2 Gait Posture Modelling Through DBN

While training a DBN, it follows two main parts, which are pre-training and fine-tune. The pre-training is based on Restricted Boltzmann Machine (RBM). Once the RBM network is pre-trained, weights of the networks are applied on fine-tune algorithm. RBM is very effective for unsupervised learning. As shown in Fig. 5, three hidden layers are applied here for RBM. RBM is basically applied to initialize the networks and can contribute to optimize errors. During initialization of the network, a greedy approach methodology is adopted. Once the weights of the first RBM are trained, first hidden layer weights get fixed. Then, the weights of the second RBM are trained using first hidden layer weights. Lastly, the third RBM is trained with previous two hidden layers weights.

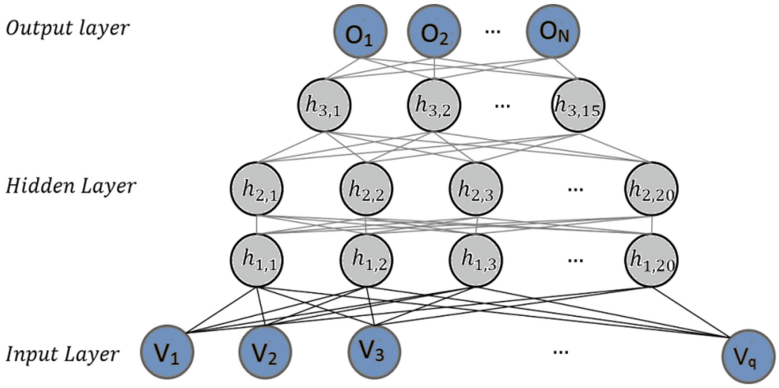


Fig. 5. Structure of a DBN.

3 Gait Posture Recognition Experimental Results

A Depth gait database had been built for two gait activities: normal and abnormal gait that consisting of 1000 images for training (i.e., 500 for each gait) and 200 images for testing (i.e., 100 for each gait). The traditional as well as proposed LDP-based shape feature extraction approaches were analyzed. The traditional shape-based recognition was applied first. In this regard, PCA, ICA, and LBP were applied to see how they perform on depth silhouette-based gait posture recognition. We started with the binary shape-based experiments first where LBP and LDP could not be applied as binary shapes represent flat binary 1 and 0 value distribution only. We started with PCA with DBN and obtained mean recognition rate of 78 %, the lowest recognition performance amongst all experiments. Later on, ICA and DBN were tried that obtained a marginal improved recognition rate of 84 %. Table 1 shows the binary shape-based gait posture recognition experiments.

After the binary silhouette-based gait posture recognition experiments, we applied depth silhouette-based approaches to see their performances. PCA with DBN was applied first that achieved mean recognition rate of 82.50 %. Then, ICA was employed

Table 1. Gait posture recognition results based on the different features from the binary silhouettes.

Feature	Gait	Recognition rate	Mean
PCA	Normal	79 %	78
	Abnormal	77	
ICA	Normal	85	84
	Abnormal	83	

that brought us a slightly improved recognition rate i.e., 88 %. Since LBP is considered to strong tool for robust local feature extraction, it was used then which achieved a better recognition rate of 91.50 %. Finally, the proposed LDP features were tried with DBN which brought the highest recognition rate of 97 %. The experimental results are shown in Table 2. Thus, our proposed LDP-DBN approach shows the superiority over other traditional approaches.

Table 2. Gait posture recognition results based on the different features from the depth silhouettes.

Feature	Gait	Recognition rate	Mean
PCA	Normal	83 %	82.50
	Abnormal	82	
ICA	Normal	89	88
	Abnormal	87	
LBP	Normal	91	91.50
	Abnormal	92	
LDP	Normal	98	97
	Abnormal	96	

4 Conclusion

In this paper, a novel deep learning-based approach has been proposed for human gait posture recognition by using depth silhouettes and LDP features in combination with DBN. LDP-based features of the depth silhouettes were investigated here and obtained superior recognition results over the traditional silhouette-based approaches such as PCA, ICA, and LBP. The proposed approach system can be employed successfully in smart homes for better human computer interaction.

Acknowledgement. This work was supported by the Samsung Research Fund, Sungkyunkwan University, 2015.

References

1. Lee, D., Lee, S., Kim, J.T., Kim, S.: A lifecycle health performance tree for sustainable healthy buildings. Indoor Built Environ. **21**(1), 16–27 (2012)
2. Zheng, Q., Lee, D., Lee, S., Kim, J.T., Kim, S.: A health performance evaluation model of apartment building indoor air quality. Indoor Built Environ. **20**(1), 26–35 (2011)
3. Boyce, P.R.: Review: the impact of light in buildings on human health. Indoor Built Environ. **19**(1), 8–20 (2010)
4. Chan, M., Estève, D., Escriba, C., Campo, E.: A review of smart homes-Present state and future challenges. Comput. Methods Programs Biomed. **91**(1), 55–81 (2008)
5. Kim, M.J., Oh, M.W., Cho, M.E., Lee, H., Kim, J.T.: A critical review of user studies on healthy smart homes. Indoor Built Environ. **22**(1), 260–270 (2013)
6. Kale, A., Sundaresan, A., Rajagopalan, A., Cuntoor, N., Chowdhury, A.R., Kruger, V., Chellappa, R.: Identification of humans using gait. IEEE Trans. Image Process. **13**, 1163–1173 (2004)
7. Robertson, N., Reid, I.: A general method for human activity recognition in video. Comput. Vis. Image Underst. **104**(2), 232–248 (2006)
8. Agarwal, A., Triggs, B.: Recovering 3D human pose from monocular images. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 44–58 (2006)
9. Howe, N.R.: Silhouette lookup for monocular 3D pose tracking. Image Vis. Comput. **25**, 331–341 (2007)
10. Niu, F., Abdel-Mottaleb, M.: View-invariant human activity recognition based on shape and motion features. In: Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering, Washington, DC, USA, pp. 546–556, 13–15 December 2004
11. Niu, F., Abdel-Mottaleb, M.: DBN-based segmentation and recognition of human activities from video sequences. In: Proceedings of IEEE International Conference on Multimedia & Expo, Amsterdam, The Netherlands, pp. 804–807, 6–8 July 2005
12. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time-images using hidden markov model. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, 15–18 June, Champaign, Illinois, USA, pp. 379–385 (1992)
13. Munoz-Salinas, R., Medina-Carnicer, R., Madrid-Cuevas, F.J., Carmona-Poyato, A.: Depth silhouettes for gesture recognition. Pattern Recogn. Lett. **29**, 319–329 (2008)
14. Feng, X., Pietikainen, M., Hadid, A.: Facial expression recognition with local binary patterns and linear programming. Pattern Recogn. Image Anal. **15**(2), 546–548 (2005)
15. Ojala, T., Pietikainen, M.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**, 971–987 (2002)
16. 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)
17. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on local binary patterns: a comprehensive study. Image Vis. Comput. **27**, 803–816 (2009)
18. Jabid, T., Kabir, M.H., Chae, O.: Local directional pattern (LDP) a robust image descriptor for object recognition. In: Proceedings of the IEEE Advanced Video and Signal Based Surveillance (AVSS), August 29 – September 1, Boston, USA, pp. 482–487 (2010)
19. Minsky, M., Papert, S.: Perceptrons. An Introduction to Computational Geometry, vol. 165 (3895), pp. 780–782. MIT Press, Cambridge (1969)
20. Hinton, G.E., Osindero, S., Teh, Y.-W.: A fast learning algorithm for deep belief nets. Neural Comput. **18**, 1527–1554 (2006)

Implementation of an Image Restoration with Block Iteration Method for Spatially Variant Blur Models

Ji Yeon Lee[1(✉)], Kuk Won Ko[2], and Sangjoon Lee[2]

[1] Automation and Energy Technology Institute, Sunmoon University, Asan, South Korea
jylee2930@gmail.com
[2] Division of Mechanics and ICT Convergence Engineering, Sunmoon University, Asan, South Korea
{kuks2309,mcp94lee}@sunmoon.ac.kr

Abstract. In image restoration, for spatially variant blur, block-wise iterative method have been proposed. Block-wise iterative method is based on the assumption that the blur approximates to spatially invariant in a small region of the blurred images. These approaches show approximation errors and block artifacts. In this work, we suggest to use block iterative method without approximates to spatially invariant blur considering arbitrary shaped blocks related to shapes of blur models. We can reduce the approximation errors and block artifacts with high speed of iteration convergence. This proposed image restoration method can reduce the computational efforts and is useful to embedded system such as mobile phone and embedded vision system.

Keywords: Spatially variant blur · Block selection · Block iteration method

1 Introduction

Spatially variant deblurring problem is computationally infeasible, because fast Fourier transform based method cannot be used [1]. To address this difficulty, block-wise iterative method have been proposed [2]. Sectioning method decompose the blurred image into several blocks equivalently, and deblur each sub-images separately using its approximate spatially invariant blur, and then combines those block-images to form the final image [3]. This approaches, however, suffer from approximation errors and block artifacts. In this work, for the spatially variant deblurring, we suggest to use the spatially-variant blur model itself in a small region. We propose the Separate block iteration method (SBI) and the Interlaced block iteration method (IBI). These two methods are based on Richardson-Lucy method. The Separated block iteration method (SBI) is same to the sectioning method [3], but the separated block iteration method (SBI) use the spatially-variant blur model itself.

We propose to use the arbitrary block shapes which is related to the blur models. For example, we consider the rectangular blocks for Gaussian blur and the diagonal blocks for Motion blur. For accelerating speed of convergence, Interlaced block

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_65

iteration method (IBI) uses a small portion of observed data sequentially in each iteration. As iteration goes, we choose whole data evenly. In simulation study, as spatially variant blur model, we consider the Gaussian blur model and the diagonal blur model. Our proposed method have many advantages. Real-time image reconstruction using small amount of pixels, even more precise images without development of devices is possible.

2 Definition and Background

2.1 Image Deblurring Problem

We model the image restoration problems as:

$$g = Pf + n \tag{1}$$

Where g is the observed image, P is the matrix that determined the blurring process, f is the original true image, and n is the Gaussian noise. We assume that f is defined on Ω which is the set of image pixels, and g is defined on Λ, respectively. We also assume following conditions for the blurring matrix P:

$$P_{b,v} \geq 0,\ P_{b,b} > 0,\ \Omega = \{v \,|\, P_{b,v} > 0,\ b \in \Lambda\} \tag{2}$$

For spatially invariant blur models in (1), the multiplication of the matrix P and the true image f, Pf can be expressed by truncated convolution with a point spread function (PSF) k:

$$(Pf)_b = \sum_{v \in \Omega} P_{b,v} f_v = \sum_{v \in (b - S_k)} k_{b-v} f_v \tag{3}$$

where $S_k = \{u \,|\, k_u \neq 0\}$, the support k. The PSE k is nonnegative, its components have sum 1, and the point $(0, 0) \in S_k$.

2.2 Richardson-Lucy Iteration Method

The Richardson-Lucy iteration takes:

$$\hat{f}^{n+1} = \hat{f}^n \cdot * P^t \left(g \cdot / P \hat{f}^n \right) \cdot / P^t 1 \tag{4}$$

Where $\cdot *$ and $\cdot /$ are pixel-by-pixel multiplication and division between two images. 1 is the all one image.

2.3 Sectioning Method

The sectioning method decompose Λ to t blocks Λ_i, $i = 1, 2, \ldots, t$ [3]. Then deblurring problem $g = Pf$ (for the simplicity of the presentation, we omitted the noise term n in (1) into several sub-deblurring problems

$$g^{[i]} = P^{[i]} f^{[i]},\ i = 1, 2, \cdots, t \tag{5}$$

And then, it approximates the matrix $P^{[i]}$ on Λ_i by a PSF $k^{[i]}$

$$P^{[i]} f^{[i]} \approx k^{[i]} * f^{[i]} \tag{6}$$

where Ω_i are domains on which solutions $f^{[i]}$ of approximated sub-deblurring problem (6) are defined and $P^{[i]}$ is the restriction of matrix P on $\Lambda_i \times \Omega_i$.

3 Proposed Method

3.1 Spatially Variant Blur

For the blurring matrix P, we consider the spatially variant Gaussian blur, which is designed as follows:

$$P_{b,v} = \gamma_b \exp\left(-\frac{1}{2\sigma^2} \left(\alpha_v (b_1 - v_1)^2 + \beta_v (b_2 - v_2)^2 \right) \right) \tag{7}$$

$$\begin{aligned} \alpha_v &= \left(c(|v_1 - v_2| + |v_1|) + 1 \right)^{-1} \\ \beta_v &= \left(c(|v_1 - v_2| + |v_2|) + 1 \right)^{-1} \end{aligned} \tag{8}$$

Here the parameter c in (9) is the degree of spatial dependence of the blurring model and the constant γ_b in (8) is introduced to make $\sum_{v \in \Omega} P_{b,v} = 1$ in (3) for each $b \in \Lambda$. For the P, we consider the spatially variant Motion blur, which is designed as follows:

$$P_{b,v} = \begin{cases} \tilde{\gamma}_b (1 + i + j + |i + j|)^{-1} & \text{if } 0 \le i, j < n_v \\ & \text{and } |i - j| \le m_v \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$\begin{aligned} i &= b_1 - v_1, j = b_2 - v_2 \\ m_v &= [n_v / v], n_v = \delta + [(|v_1| + |v_2|)/\mu] \end{aligned} \tag{10}$$

Here, parameter δ, μ, v are introduced to control the size and the degree of the motion blur. The notation $[x]$ denotes the largest integer that is smaller than or equal to x. The constant $\tilde{\gamma}_b$ is introduced to make $\sum_{v \in \Omega} P_{b,v} = 1$ in (3) for each $b \in \Lambda$.

3.2 Separate Block Iteration Method

The Separated block iteration method (SBI) is similar to the sectioning method [4], but the separated block iteration method (SBI) use the spatially-variant blur model itself (Fig. 1).

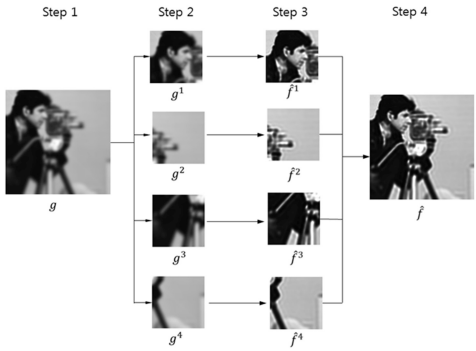


Fig. 1. The process of the separate block iteration method

3.3 Interlaced Block Iteration Method

Interlaced block iteration method (IBI) uses the spatially variant blur model itself, and uses blocks that consist of scattered pixels(down-sampled pixels) and immediately uses iterates from the current block as the starting image for the next iteration [5] (Fig. 2).

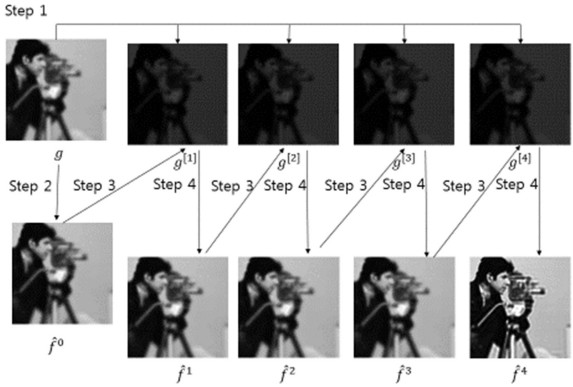


Fig. 2. The process of the interlaced block iteration method

3.4 Block Selection

We use arbitrary shaped blocks to increase flexibility of selection of block [5]. 4×4 rectangular down-sampled blocks are defined by

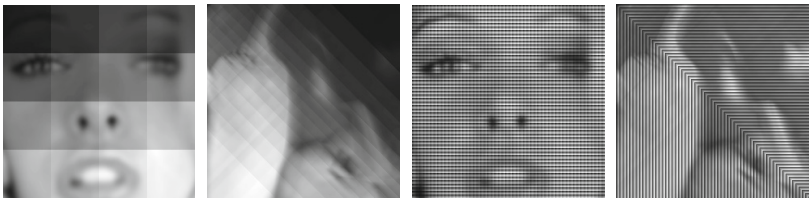
$$\Lambda_{4y+z+1} = \{(i_1, i_2) \in \Lambda \,|\, i_1 \in 4z + y,\ i_2 \in 4z + x\} \text{ for } x, y = 0, 1, 2, 3 \qquad (11)$$

Where $4z$ is the integer subset formed by multiples of 4. Diagonally down-sampled blocks of size 8 are defined by

$$\Lambda_{i+1} = \{(i_1, i_2) \in \Lambda \,|\, i_1 + i_2 \in 8z + t\} \text{ for } t = 0, 1, 2, 3, \ldots, 7 \qquad (12)$$

4 Simulation and Results

To compare with proposed method, we simulate sectioning method, and interpolation method. We use the spatially variant Gaussian blur model with $\sigma = 0.5, c = 0.2$, and Motion blur model with $a = 7, b = 20, c = 5$ in (10) and (11). We measure the time to construct spatially variant blur PSFs, t_p. t_{RL} is the time for one round iteration of RL. t^*_{RL} is an average of t_{RL}. MSE is the mean square error (Figs. 3, 4, 5 and 6).

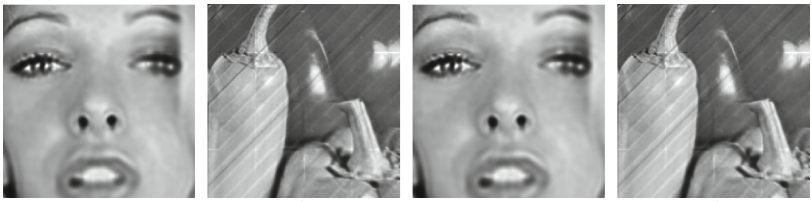


(a) (b) (c) (d)

Fig. 3. (a) 4×4 rectangular (b) 16 diagonal (c) 4×4 rectangular down-sampled (d) 4 diagonally down-sampled blocks

(a) (b) (c) (d)

Fig. 4. Test images (a) Face (b) Pepper Blurred images (c) by spatially variant Gaussian blur (d) spatially variant motion blur



(a) (b) (c) (d)

Fig. 5. Deblurred images (a) (b) by sectioning method (c) (d) by interpolation method

(a) (b) (c) (d)

Fig. 6. Deblurred images (a) (b) by separated BI method (c) (d) by Interlaced BI method

Table 1. Section: Sectioning method, Interpol: Interpolation method, SBI: Separated block iteration method, IBI: Interlace block iteration method.

		Section	Interpol	SBI	IBI
t_p	Face	1.19	1.19	18.93	18.93
	Pepper	1.18	1.18	1.25	1.25
$t_{RL}(t^*_{RL})$	Face	0.29	0.29	0.15	0.00095
	Pepper	0.29	0.29	0.034	0.0361
MSE	Face	7.22	7.20	7.05	7.07
	Pepper	6.31	6.26	5.01	4.63

5 Conclusion and Further Study

Simulation results show that, through the comparison to t_p, $t_{RL}(t^*_{RL})$, MSE, the separated block iteration (SBI) and the interlaced block iteration (IBI) can reduce computational burden by high speed of accelerating iterations. We can also avoid approximation errors and block artifacts. This proposed method can be applied to 3D X-ray microscopy, PET Image Reconstruction, digital camera and smart phone with a new function. Our next goal is to deblur the images from unknown blurring models (Table 1).

Acknowledgment. This research was financially supported by the 2014 Ministry of Education (MOE) and National Research Foundation of Korea (NRF) through the Human Resource Training Project for Regional Innovation and Creativity (NRF-2014H1C1A1066998).

References

1. Berisha, S., Nagy, J.G.: Iterative methods for image restoration. Sig. Process. Image Video Process. Anal. Hardware Audio Acoust. Speech Process. **4**, 193 (2013). Academic Press Library
2. Robbins, G.M., Hung, T.S.: Inverse filtering for linear shift-variant imaging systems. Proc. IEEE **60**, 862–872 (1972)
3. Trussell, H.J., Hunt, B.R.: Image restoration of space-variant blurs by sectional methods. IEEE Trans. Acoust. Speech Sig. Process. **26**, 608–609 (1978)
4. Nagy, J.G., O'Leary, D.P.: Fast iterative restoration with a spatially-varying PSF. In: Proceedings of SPIE. Advanced Signal Processing: Algorithms, Architectures, and Implementations VII, vol. 3162, p. 388 (1997)
5. Lee, N.-Y.: Block iterative Richardson-Lucy methods for image deblurring. EURASIP J. Image Video Process. **2015**, 14 (2015). doi:10.1186/s13640-015-0069-2

Analysis of Hard Shadow Anti-aliasing

Hua Li[1], Yuling Cao[1], and Xin Feng[1,2(✉)]

[1] School of Computer Science and Technology,
Changchun University of Science and Technology, Changchun 130022, China
custlihua@gmail.com
[2] National Local Combined Engineering Research Center of Special Film
Technology and Equipment, Changchun 130000, China

Abstract. Shadows are essential elements in computer-generated scenes. Shadow mapping is one of the most popular techniques for rendering shadows in real-time applications which is widely supported in current graphics hardware with its flexibility independence to the complexity of the scene. However, the standard shadow maps suffer from aliasing artifacts due to finite resolution and bias problems. In this paper, we survey the most promising research results that have been achieved recently and conclude by discussing some of the most important challenges of shadow aliasing.

Keywords: Shadow mapping · Anti-aliasing · Real-time · Resolution

1 Introduction

Shadows are important results of the light transport in a virtual scene. They give visual cues and geometric relationships of objects each other. Shadow mapping is the most popular algorithm in real-time applications. However, shadow mapping suffers from serious artifacts due to the mismatching of different sample distributions. This paper discusses the issues of aliasing of hard shadows generated by shadow mapping and the most promising research results have been achieved especially for the past five years. We draw conclusions with different techniques.

Shadow mapping is an image-based algorithm; it can be achieved within two passes (see Fig. 1). In the first pass, shadow mapping renders the scene from the light's point of view and store visible depth value as textures in z-buffer (shadow map). In the second pass, the scene is rendered from camera's (eye's) view. Each pixel is transformed into light space to compare its z value with the corresponding depth value stored in the shadow map to determine a view sample being shadowed or illuminated.

2 Analysis of Aliasing

The straightforward understanding aliasing occurred in shadow mapping is the mismatching of pixels and texels. An insightful observation for aliasing is given with a mismatching ratio of the beam footprint widths [1] (Eq. 1 and Fig. 1):

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_66

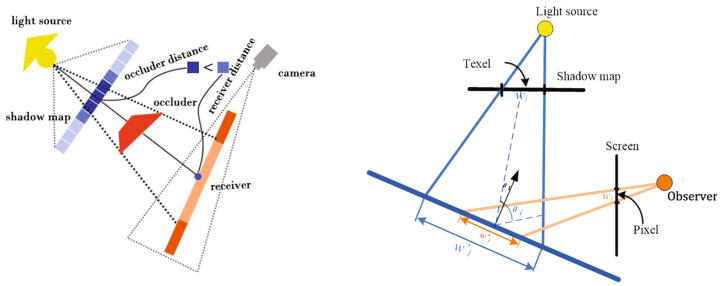


Fig. 1. The principle of shadow mapping (left), shadow map aliasing (right).

$$m = \frac{w_l^{'}}{w_i^{'}} \approx \frac{w_l}{w_i} \frac{\cos\theta_i}{\cos\theta_l} \tag{1}$$

Where w_i and w_i are the widths of the image and light beams at the point of intersection. θ_i and θ_l are the angles between the surface normal and the beam directions. w_l/w_i is referred to as perspective aliasing and $\cos\theta_i/\cos\theta_l$ is referred to as projection aliasing. Therefore, perspective aliasing depends solely on the relative positions of the light and camera which is independent of the scene geometry and perspective aliasing vanishes when the beam widths are the same, i.e. $w_i = w_l$. For projection aliasing, it depends on the surface orientation to the beam directions only if $\theta_i = \theta_l$, where the surface is parallel or perpendicular to the half-way vector between the beam directions.

3 Anti-aliasing Technology

Good reviews can be found for shadow generating technology in the past years. In 2011, Scherzer [2] analyzed the technique of hard shadow generation and the main types of error occurred in hard shadow technique were presented. They are basically can be summarized in two reasons: sampling error and reconstruction error. Many methods and strategies reduce the sampling error, such as focusing methods, which reduce the shadow map space of the invisible scene parts; warping methods which make sure that higher sampling densities for the near viewpoint and lower sampling densities for the far viewpoint; partitioning methods using multiple shadow maps try to approximate the ideal sample distribution; irregular sampling methods which try to create shadow map samples exactly in those positions that will be required later on; finally temporal re-projection methods, which reuse samples from previous frames through re-projection. Kolivand [3] also gives a review about shadow techniques. In the paper, we will analyze the anti-aliasing technology of the newest method proposed mostly in the past five years through another view of the technology.

3.1 Adaptive Sampling Anti-aliasing

Adaptive sampling method traces the edges among the previous rays using additional rays [4], which is suitable for real time-moving shadow detection. This method presented a method improving shadow detection accuracy by adaptive threshold estimator. But it suffers from the huge number of required rays. Gumbau [5] proposed a statistical filtering method that approximated the cumulative distribution function (CDF) of depth values by a Gaussian CDF, which significantly reduced "light leaks" and has good performance.

Barák T. et al. proposed a new method called virtual point light (VPL) [6] based on Metropolis-Hastings sampling. The familiar problem about M-H sampling is that the samples are not followed by target distribution. Thus, it may lead to the so-called start-up bias of the estimator. See Fig. 2, a delayed shading pipeline should be used in the first stage and a layered frame buffer representation of the camera view should be created. For each primary light source, the algorithm creates reflective shadow map (RSM) for them. The method is fast requiring with no additional data storage and significantly reduces temporal flickering.

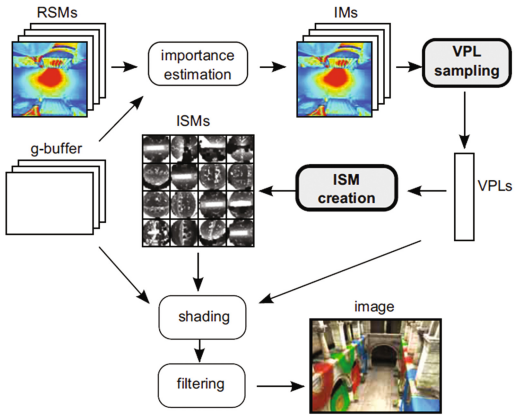


Fig. 2. Indirect lighting calculating based on incomplete view of adaptive shadow maps [6].

Jia *et al.* [7] proposed a method called Distorted Shadow Maps (DSMs), which shrinked the region that is out of the penumbra regions to enlarge the shadow's silhouette. It also adopted redistribution mechanism to solve the problem of aliasing. Thus, the method transformed more texels by using geometric silhouettes and could indicate more samples.

Wang *et al.* [8] settled the problem of aliasing using shadow geometry maps, which reduced the discrete approximation of depth values. A super-sampling filter stragety and a geometry-aware reconstruction kernel are combined to this algorithm; it enhanced the anti-aliasing consequently.

3.2 Perspective Space Anti-aliasing

Stamminger *et al.* [9] and Wimmer *et al.* [10] proposed shadow mapping technologies respectively. Perspective shadow map mainly generated the shadow map in the perspective space, and it could replace the standard shadow mapping. LiPSM used the perspective transform in light space to solve the perspective aliasing.

The idea of the perspective shadow map transforms the scene to post-perspective space. In the world space, the scene is illuminated by a vertical directional light source as shown in Fig. 3. The directional light sources can be seen as the infinite point lights. The shadow map is an orthogonal view which is from the top onto the unit cube in the post-perspective space. If the shadow map pixels which projected on the ground have the same size as the image in the scene, the perspective aliasing is avoided.

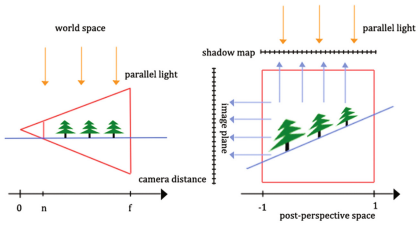


Fig. 3. The transformation of perspective shadow map.

The method of PSM can be used for point rendering, the method compared to the uniform shadow map allowed to generate the shadow map. However, the defects of the PSM are obvious: as the lights were transformed in the perspective space and the types of lights changed, the perspective space was not intuitive by transforming in the special view. The mapping after perspective space has a singularity; it could result in the shadow casters and the opposite singularity.

LiPSM based on the perspective projection could process all the lights as directional lights and the perspective projection has no relevant singularity. It is applied widely by computing the convex hull which combines with the view frustum and the directional light. The perspective frustum P which is parallel to the shadow map enclosed the convex hull (see Fig. 4). P is mainly used in the shadow map generated rendering.

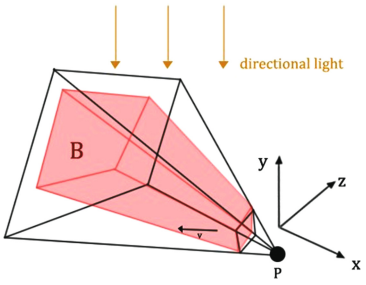


Fig. 4. The perspective frustum of P in 3D.

LiPSM is an effective shadow mapping method, which integrates the benefits of perspective and uniform shadow maps. LiPSM can provide higher shadow quality and eliminate the shortcomings of perspective shadow maps. Therefore, the method is as fast as standard shadow maps and easy to come true. Kolic *et al.* [11] presents a method called Camera Space Shadow Maps (CSSM) which is a fast non-linear projection technique for shadow map stretching that enables complete utilization of the shadow map by eliminating wastage. As a result, it reduces both aliasing and wastage by pushing the camera backward in the space of camera and handling the dueling frustum. It mainly produces high quality shadows and increases the X-resolution of shadow maps. CSSM is a single-pass shadow mapping technique, and it achieves better shadow quality than LiPSM.

3.3 Sub-pixel Anti-aliasing

Sub-pixel anti-aliasing is a new technique proposed in recent years. The quality of shadow can be improved by increasing the resolution of shadow map. However, it has the issue of incredible memory. Sub-Pixel Shadow Mapping [12] is proposed to avoid the perspective aliasing and projection aliasing based on reconstructable geometry shadow maps [13]. This is a great idea that reconstructs depth value of a view sample using geometry information. As for the coverage of overall triangle, the generation of the Sub-pixel antialiasing actualizes the generation of fragments by Conservative Rasterization. As seen in Fig. 5, the fragments produced for triangles that intersect the texel area are guaranteed by Conservative Rasterization.

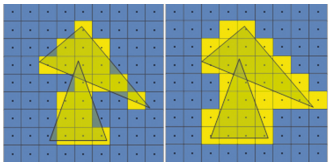


Fig. 5. a. Standard rasterization b. Conservative rasterization.

In SPSM, a compact triangle storage strategy is used in Fig. 6, which ensures only a small computational overhead than this method [13]. A shadow silhouette recovery is enforced for generating hard shadows, which is combined with precision of sub pixel and consistency of high temporal.

Wang *et al.* [8] proposed a method which could produce high quality alias-free and generate subpixel super sampling shadows. The algorithm is a new way to combine the geometry-aware reconstruction kernel with the super sampling filter, and an irregular projection scheme. As is shown in Fig. 7(a), the shadow geometry map and memory layout of the cell-primitives is listed. in Fig. 7(b), each cell related to primitives are consecutively stored in a 1D list.

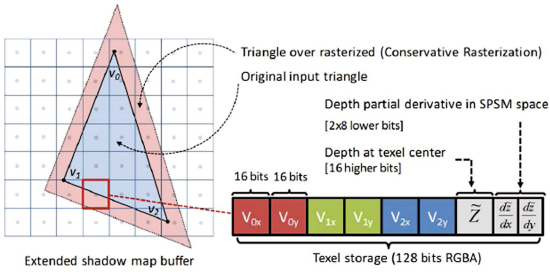


Fig. 6. Conservative rasterization in the shadow map and SPSM stored for each texel [11].

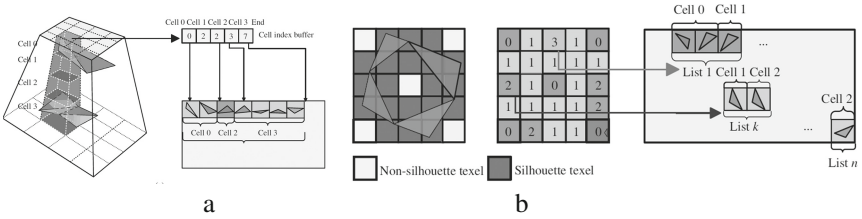


Fig. 7. Illustration of shadow geometry map and memory layout [8].

3.4 Partitioning Algorithms

For the algorithm of partitioning, the z-partitioning is used to render the scene. From the perspective of the light source, the view frustum is splitted by face partitioning. Face partitioning allows warping to be used when it could not be used otherwise leading to reduced error [1].

Lefohn *et al.* [14] proposed a method based on the adaptive partitioning which called Resolution Matched Shadow Maps (RMSMs). The idea of this method is storing data in a quadtree and generates shadows in a single step. RMSMs required a little geometry because of the approximation of sample positions.

Liang *et al.* [15] concerned with the method called Cascade Shadow Map (CSD), in which the light direction and view direction are not vertical. LiSCSM (Light Space Cascade Shadow Map) presented by them improved the method. The main idea is to create shadow maps for each layer in the light space by quitting the scene in non-intersection layers.

Adaptive Volumetric Shadow Maps (AVSM) was presented by Salvi *et al.* [16], which is built upon adaptive shadow algorithm (ASM) using deep maps. This shadow algorithm could product high quality shadows in real-time; it mainly supports dynamic volumes of media *i.e.* smoke, fur and hair.

4 Depth Bias Anti-aliasing

The shadow depth map revisited [17] records the average depth of the nearest and second nearest surface in the shadow map. But in the problem of silhouettes, false self-shadowing remains. Han *et al.* [18] presents a adaptive depth bias method to eliminate false self-shadowing. It estimates the potential shadow caster for each fragment and computes the minimal depth bias. For each pixel, a constant depth bias is used by casting curved shadows on curved surfaces before shadow testing, incorrect self-shadowing is removed by constant bias. However, it produces the self-shadowing and shadow detachment.

This method locates a potential occluder for a given fragment. As is depicted in Fig. 8 (right), $\vec{R}$ is the ray traced from the light source through the texel center C. The intersection of $\vec{R}$ and P can get the potential occluder F_2 for a given fragment F_1.

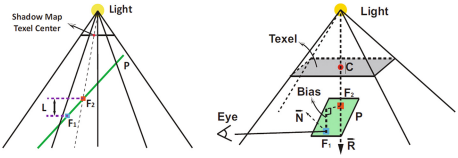


Fig. 8. False self-shadowing (left), optimal depth bias (right) computation for a traditional shadow map.

In optimal depth bias, a small epsilon value is needed to move F_1 just above its occluder F_2. The calculating formula of depth bias is Eq. 2:

$$adaptiveDepthBias = optimalDepthBias + adaptiveEpsilon \tag{2}$$

King G. [19] proposes a depth bias of slope-scale computed based on the fragment's depth slope relative to light view. Method for generating shadows in an image [20] is weighted by the angular deviation of the depth of the surface normal and the incident between the line segments. The two methods are not generated for each segment, providing the minimum deviation which is still a certain degree of false unshadowing. Compared to the method of adaptive depth bias for shadow maps, they have many errors.

5 Conclusion

In this survey, we have discussed a variety of famous shadow algorithms. An overview on each of them is presented. We give some practical hints and conclusion about advantages and techniques (Table 1). As we mentioned above, there is no perfect method can solve all the situations. What we can do is to choose the most suitable method to meet our goals.

Table 1. A summary of anti-aliasing shadow algorithms

Algorithms	Advantages	Limitations	Technique
DSM	Accurate shadows	Hard shadowing	Redistribution
SPSM	Sub-pixel shadows	Exist shadow fusion	Reconstruction
PSM	Camera capturing	Consider light position	Normalize
LiPSM	Fast rendering	Require parameterization	Adaptive partitioning
CSM	For large scenes	Redundant rendering	Z-partitioning
LiSCSM	For wide scenes	Soft shadow	Partitioning
ASM	High quality	Iterative refinement	Adaptive partitioning
AVSM	For complex shadow	For graphic hardware	Adaptive partitioning

Acknowledgements. This work was financially supported by development project of Jilin province science and technology (20140204009GX), Jilin upgrade industrial innovation special fund projects (2016C091) and major scientific and technological plan of Changchun (14KG008).

References

1. Lloyd, D.B., et al.: Warping and partitioning for low error shadow maps. In: Proceedings of the 17th Eurographics Conference on Rendering Techniques Eurographics Association, pp. 215–226 (2006)
2. Scherzer, D., Wimmer, M., Purgathofer, W.: A survey of real-time hard shadow mapping methods. Comput. Graph. Forum **30**(1), 169–186 (2011). (Blackwell Publishing Ltd.)
3. Kolivand, H., Sunar, M.S., Altameem, A., et al.: Shadow mapping algorithms: applications and limitations. Appl. Math. Inf. Sci. **9**(3), 1307 (2015)
4. Jiang, K., Li, A., Su, Y.: Adaptive shadow detection based on global texture and sampling deduction. J. Optoelectron. Laser **11**, 25 (2012)
5. Gumbau, J., Sbert, M., Szirmay-Kalos, L., et al.: Smooth shadow boundaries with exponentially warped gaussian filtering. Comput. Graph. **37**(3), 214–224 (2013)
6. Barák, T., Bittner, J., Havran, V.: Temporally coherent adaptive sampling for imperfect shadow maps. Comput. Graph. Forum **32**(4), 87–96 (2013). (Blackwell Publishing Ltd.)
7. Jia, N., Luo, D., Zhang, Y.: Distorted shadow mapping. In: Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology, pp. 209–214. ACM, New York (2013)
8. Wang, R., Wu, Y.Q., Pan, M.H., et al.: Shadow geometry maps for alias-free shadows. Sci. China Inf. Sci. **56**(6), 1–12 (2013)
9. Stamminger, M., Drettakis, G.: Perspective shadow maps. ACM Trans. Graph. (TOG) **21**(3), 557–562 (2002). (ACM)
10. Wimmer, M., Scherzer, D., Purgathofer, W.: Light space perspective shadow maps. In: 15th Symposium on Rendering Techniques (2004)
11. Kolic, I., Mihajlovic, Z.: Camera space shadow maps for large virtual environments. Virtual Reality **16**(4), 289–299 (2012)
12. Lecocq, P., Marvie, J.E., Sourimant, G., et al.: Sub-pixel shadow mapping. In: Proceedings of the 18th Meeting of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 103–110. ACM, New York (2014)

13. Dai, Q., Yang, B., Feng, J.: Reconstructable geometry shadow maps. In: Proceedings of the 2008 Symposium on Interactive 3D Graphics and Games, p. 4. ACM, Redwood City (2008)
14. Lefohn, A.E., Sengupta, S., Owens, J.D.: Resolution-matched shadow maps. ACM Trans. Graph. (TOG) **26**(4), 20 (2007)
15. Liang, X.H., Ma, S., Cen, L.X., et al.: Light space cascaded shadow maps algorithm for real time rendering. J. Comput. Sci. Technol. **26**(1), 176–186 (2011)
16. Salvi, M., Vidimče, K., Lauritzen, A., et al.: Adaptive volumetric shadow maps. Comput. Graph. Forum **29**(4), 1289–1296 (2010). (Blackwell Publishing Ltd.)
17. Woo, A.: The Shadow Depth Map Revisited. Graphics Gems III, pp. 338–342. Academic Press Professional Inc., San Diego (1992)
18. Dou, H., Yan, Y., Kerzner, E., et al.: Adaptive depth bias for shadow maps. In: Proceedings of the 18th Meeting of the ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, pp. 97–102. ACM, San Francisco (2014)
19. King, G.: Shadow mapping algorithms. GPU Jackpot Presentation, pp. 354–355 (2004)
20. Gautron, P., Marvie, J.E., Briand, G.: Method for generating shadows in an image: U.S. Patent Application 13/138, 652, pp. 3–26 (2010)

Cascade-Adaboost for Pedestrian Detection Using HOG and Combined Features

Gyujin Jang, Jinhee Park, and Moonhyun Kim(✉)

College of Information and Communication Engineering,
Sungkyunkwan University, 2066 Seobu-ro, Jangan-gu,
Suwon, Gyeonggi-do, Republic of Korea
{gjjang, joshdev, mhkim}@skku.edu

Abstract. Over the recent years, pedestrian detection beings in a video surveillance system is attracting more attention due to its wide range of applications. In this paper, we propose an efficient two-phase pedestrian detector using HOG and combined features. The detector finds pedestrian candidate regions with a cascade-adaboost on HOG features. It then verifies each candidate using a combined features, which is local (SURF) and global features (RGB histogram), and then a classification based on MLP. It obtains a better detection rate and false-positive rate. The pedestrian detection system experimented with PETS 2009 dataset proves the effectiveness of our detection model.

Keywords: Pedestrian detection · Classifiers · Cascade · Adaboost · MLP · HOG · Global features · Local features · Color histogram

1 Introduction

In recent years, pedestrian detection is an important area of object recognition, which has been widely used in various domains such as video surveillance, advanced driver assistance systems, and smart cameras [1, 2]. It is a challenging task because of many difficulties. The appearances of people are different; the illumination is changing. Pedestrian detection consists with two important parts: feature extraction and detector classifier. Pedestrian detection has three steps. Firstly, extracting characteristics of pedestrian, such as the contour features, color features of the pedestrian or texture features of the pedestrian etc., then training a classifier, finally gained the pedestrian detector, it can be used for pedestrian detection in images and videos. The pedestrian detection in image will provide the location information and the identification of the pedestrian.

The most critical part in pedestrian detection is how to extract effective pedestrian features on which the overall performance of detection depends. The detection performance of combining local and global features is much better than the single feature [3]. Several pedestrian detection approaches have been proposed in the literature, various features have been studied. These features can be computed from low-level information

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2014R1A2A1A11053902).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_67

such as edge, texture, color, or motion which are global features. There is much ongoing research in exploring a feature for pedestrian detection. The local features include haar features [4], Scale invariant feature transform (SIFT) [5], speeded-up robust feature (SURF) [6], histogram of oriented gradient (HOG) [7], local binary pattern (LBP) features [8], local self-similarity (LSS) features [9], covariance features [10], multi-scale orientation (MSO) features [11], shapelet features [12] etc. In [13], the color feature CHOG (described above) was combined with HOG. So we suggest to combine local and global features as accumulated (each) color channel and SURF with HOG.

The two-stage classifier proposed by Guo et al. [14], can further reduce false positive rates and this system has better performance than these single-stage algorithms. However, the detection rates cannot be further increased and maintained at a certain level as can these single-stage algorithms. Our goal is to lower the number of false positives without too much penalty in the detection rate. So, MLP and cascade-adaboost are selected as dominant classifiers. These two algorithms can simultaneously deal with the normalized candidates extracted by multiscale sliding windows, which can guarantee the detecting efficiency of the proposed system. It's expected to find a combination feature with the highest detecting precision.

This paper is organized as follows: related work will be briefly introduced in Sect. 1; in Sect. 2, we describe our proposed system in detail; and experiment results will be shown in Sect. 3. Lastly the conclusion and future work is discussed in Sect. 4.

2 Overview of Proposed Pedestrian Detection Model

In this paper, a novel two-phase detecting system is proposed based on a cascade-adaboost detector and MLP. These processing results of the cascade-adaboost detector and MLP are fused together, as the final evaluation criteria of whether these candidates are pedestrians or not. The Architecture of our system is illustrated in Fig. 1.

We select carefully the training method used in each stage, particularly the first stage, since it is applied to train the whole image features. Consequently, adaboost is adopted in the first stage to learn HOG feature because it has proved a good

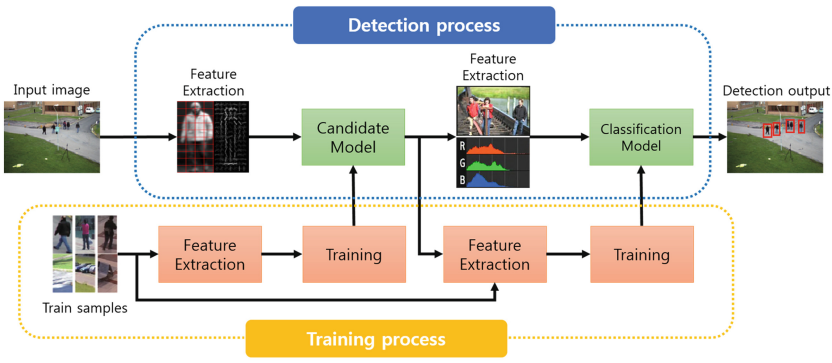


Fig. 1. Architecture design of our pedestrian detection system

performance in pedestrian detection. In the second stage, MLP is selected to train HOG features of the candidate regions supplied by the first stage. The classification methods used in this paper are briefly explained in the following two Sects. 2.1 and 2.2.

2.1 Extraction of Pedestrian Candidate Regions

In the first phase, the cascaded classifiers are trained to extract pedestrian candidates. The training of cascaded classifiers includes local feature and learning via adaboost algorithm. Feature extraction is a type of dimensionality reduction that efficiently represents the ROI region of an image in the fields of object detection and pattern recognition algorithms. These features are extracted as a compact feature vector, for subsequent processing. Therefore, effective image feature extraction is rather important, which concerns detection accuracy.

This paper uses HOG, one of the way to extract feature which has robust feature to the contrast caused by irregularly change of light and noise in object detection.

We generate histogram by accumulating the direction of edge's incident frequency in topical area. HOG features are adopted to enhance the pedestrian detection performance of the proposed system. In our system, the parameters for the HOG feature extraction applied to use cascaded classifiers. In order to calculate HOG feature Vector, we need to determine the slope value m and direction θ from pixel values inside of the area by 2 formulas below.

$$\begin{aligned} m(x_i, y_i) &= \sqrt{\partial_x(x_i, y_i)^2 + \partial_y(x_i, y_i)^2} \\ \theta(x_i, y_i) &= \arctan\left(\frac{\partial_y(x_i, y_i)}{\partial_x(x_i, y_i)}\right) \end{aligned} \tag{1}$$

For our system, the normalized candidates are divided into 16 × 16 pixel blocks; each cells size of 8 × 8 pixels; linear gradient voting into 9 orientation bins in (0°, 360°). It gets histogram about repetition rate in 9 bins. Figure 2 shows a consequence which is got HOG feature by pedestrian image.

To further reduce the false alarm rate, our system make use of a cascade-adaboost classifier, the architecture of which is shown in Fig. 2, where each classifier is adaboost classifier. The cascade classifier is designed to locate pedestrian candidates which is

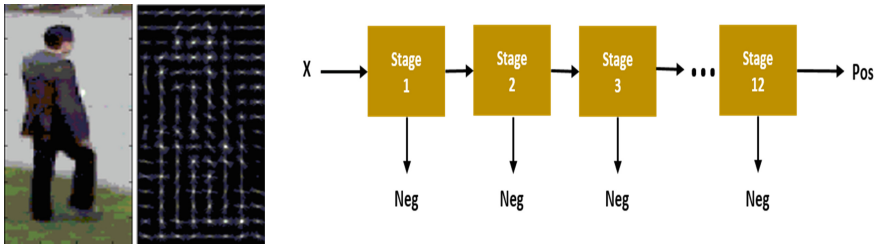


Fig. 2. Hog Feature (left) and design of cascade stages (right)

consist of 12 stages as less than 100 weak classifiers for each stage. So we utilize this classifier to generate the pedestrian candidate regions in the scene.

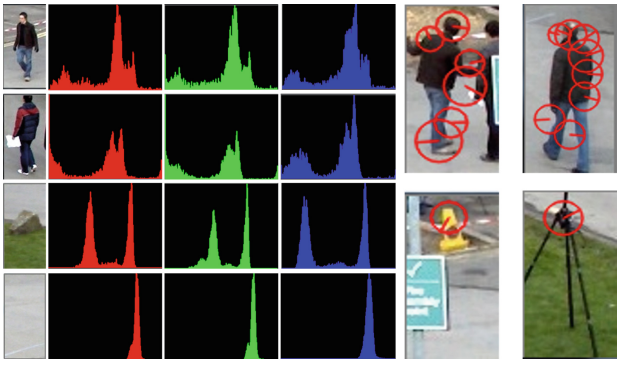
2.2 Pedestrian Classification Using Combine Features with MLP

In this phase, classifier is constituted by using local and global features. SURF (Speeded-Up Robust Features) algorithm was proposed by Bay et al. [7]. SURF is one of the most robust local invariant feature descriptors. SURF is implemented mainly for gray images. However, color presents important information in the object description. Literature [15] as the first one proposed the application of the color information of pedestrians in pedestrian detection and got a good performance, it is an innovation character of pedestrian detection. The color information is a simple low level and an information that could get through quick calculation. It creates histogram about R, G and B channels and find out classification patterns between pedestrians and none pedestrians. For extracting efficient features in color histogram, it calculates accumulate frequency by histogram of each RGB channel. $f_x(j)$ means the number which is counted j level in candidate area.

$$G(i) = \sum_{i=0}^{i} f_x(j), \quad 0 \leq i \leq 255 \tag{2}$$

In this experiment, it sets $i = 40$ for initial value. Figure 3 is color histograms using R, G, and B channels and SURF between pedestrians and none pedestrians. Figure 3 is shows the features extracted from the pedestrians and non-pedestrians.

We provide cumulative (each) color values and the number of SURF point, as the input data of MLP. Our MLP is organized of three types of layers as it follows: (i) the input layer with 4 nodes, (ii) the hidden layer with 10 nodes and the output layer with 1 node.



(a) (b)

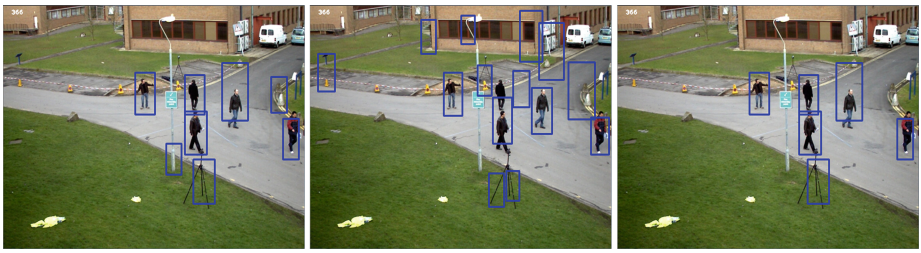
Fig. 3. The results of color histogram (a) and SURF points (b) of pedestrian and non-pedestrian

3 Experiments

Our pedestrian detection system evaluate the performance using crowd scene of sequences the PETS 2009 dataset which is composed of a total of 754 frames [16]. And we use about 2000 positive samples (normalize to size 64*128) and more than 3500 negative samples as our training dataset with INRIA and PETS 2009. The ratio is 60 % and 40 % respectively. We test our detection system on the test dataset, and compared our results with the original HOG/cascade-adaboost method and the Haar/cascade-adaboost method. Table 1 shows the performance of each model. Figure 4 present the result of pedestrian detection.

Table 1. Detection result of the three methods.

Model	Detection rate
Haar/adaboost (stage 12)	74.02 %
HOG/adaboost (stage 12)	82.45 %
Proposed system	89.3 %



(a) (b) (c)

Fig. 4. The results of pedestrian detection in test scenes. (a) HOG/adaboost (b) Haar/adaboost (c) proposed system

4 Conclusion

Recently, detection technique has been a focus of research in intelligent surveillance system especially for pedestrian detection. We proposed the method for efficient detecting in outdoor scenes. Proposed system detects pedestrian candidate region by detecting HOG features from input video and generating pedestrian model through cascade-adaboost. Then, we extract color information and SURF points from detected candidate regions and judge which is pedestrian or not from the crowd scene by MLP.

In this paper, we presented two-phase pedestrian detecting system based on a cascade-adaboost detector and MLP. We have presented the results of the proposed pedestrian detector on the PETS 2009 dataset. We have shown its effectiveness in accurately detecting multiple pedestrian.

In the future, more pedestrian detection algorithms should be studied for the enhancement of accuracy in surveillance scenes.

References

1. Viola, P., Jones, M., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: ICCV (2003)
2. Gerónimo, D., López, A.M., Sappa, A.D., Graf, T.: Survey of pedestrian detection for advanced driver assistance systems. IEEE Trans. Pattern Anal. Mach. Intell. **32**(7), 1239–1258 (2010)
3. Zhao, Z.-Q., Huang, D.S., Sun, B.-Y.: Human face recognition based on multiple features using neural networks committee. Pattern Recogn. Lett. **25**(12), 1351–1358 (2004)
4. Viola, P., Jones, M.J., Snow, D.: Detecting pedestrians using patterns of motion and appearance. In: Proceedings of the 9th IEEE Conference on Computer Vision, pp. 734–741 (2003)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
6. Bay, H., Ess, A., Tuytelaars, T., Van, L.: Speeded-up robust features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008)
7. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893 (2005)
8. 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. (2002)
9. Shechtman, E., Irani, M.: Matching local self-similarities across images and videos. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
10. Tuzel, O., Porikli, F., Meer, P.: Human detection via classification on riemannian manifolds. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
11. Ye, Q., Jiao, J., Zhang, B.: Fast pedestrian detection with multi-scale orientation features and two-stage classifiers. In: Proceedings of the 17th IEEE International Conference on Image Processing (ICIP), pp. 881–884 (2010)
12. Sabzmeydani, P., Mori, G.: Detecting pedestrians by learning shapelet features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)
13. Ott, P., Everingham, M.: Implicit color segmentation features for pedestrian and object detection. In: Proceedings of the International Conference on Computer Vision, pp. 723–730 (2009)
14. Guo, L., Ge, P.-S., Zhang, M.-H., Li, L.-H., Zhao, Y.-B.: Pedestrian detection for intelligent transportation systems combining AdaBoost algorithm and support vector machine. Expert Syst. Appl. **39**(4), 4274–4286 (2012)
15. Stefan, W.: New features and insights for pedestrian detection. In: 2010 IEEE Conference on Schiele in Computer Vision and Pattern Recognition, pp. 1030–1037 (2010)
16. 11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, June 2009. http://pets2009.net

Portable Hypervisor Design for Commercial 64-Bit Android Devices Supporting 32-Bit Compatible Mode

Kangho Kim(✉), Kwangwon Koh, Seunghyub Jeon, and Sungin Jung

ETRI, Daejeon, South Korea
{khk, kwangwon.koh, shjeon00, sijung}@etri.re.kr

Abstract. We present a hypervisor design that can be applied to any commercial 64-bit Android devices without support of set makers. We achieved the portability by using pure software virtualization while preserving high performance. The contribution of the design is to put the guest OS and the hypervisor together into a single address space which results in avoiding the address space compression problem and reducing major virtualization costs, using 32-bit compatible mode. The design using the single address space makes the hypervisor simple and run fast even with pure software technologies. Prototypical implementation of the design is composed of one kernel module and one user-level program managing virtual machines for Android OS. We have evaluated our design on a commercial mobile phone, Nexus 6P. Since any Android device allows inserting kernel modules and installing user programs on it, we think that our hypervisor can be utilized on any 64-bit ARM-based mobile phones.

Keywords: Hypervisor · Virtual machine · Mobile phone · ARM · KVM

1 Introduction

Carriers and third-party mobile service providers aspire to secure their own space inside mobile phones to execute their core logics as well as to save data only for their own services. They expect that the space must be as independent as possible from the phone makers. They want to achieve their requirement without help of the phone makers. If reserving the space needs to modify the bootloader of phones, they cannot help asking the phone maker to add some functionalities into the bootloader.

The most common way to occupy their own independent space inside a real machine is to create a virtual machine (VM). We are able to put one's own data, service programs, and an OS in the VM and keep them separate from the environment the phone makers provide. We can create VMs with QEMU or KVM that are open source hypervisors freely available. QEMU without help of KVM is a hypervisor fully emulating processor instructions and peripheral devices, so that it may not meet performance requirements. KVM using hardware virtualization extension shows higher performance than QEMU. However, it depends on the bootloader because it expects the bootloader to set the processor to hypervisor mode (EL2) prior to kernel booting.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_68

As far as we know, Samsung Galaxy S6 bootloader sets processor's mode to kernel mode (EL1) and transfers control to the Android kernel. The kernel disables KVM/ARM module after detecting it is running in the kernel mode [1, 2].

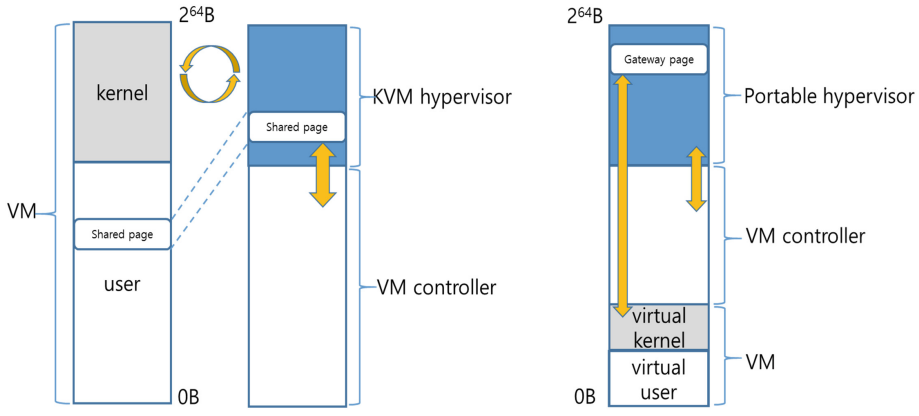
We present a portable hypervisor that provides the VM with 32-bit address space and expect its performance is as high as KVM, without cooperation of the bootloader. We assume that core part of both data and service program is not so large that 32-bit address space would be enough to accommodate those data and service in it. The 32-bit restriction help us to achieve efficiency of the hypervisor and keep it small even without using ARM virtualization extension.

2 Hypervisor Design

We also assume that ARM virtualization extension is not available, so that we have no choice but to resort to pure software hypervisor technologies. We prefer to full virtualization, which doesn't require any source code change for the guest OS(OS running inside a VM). The popular design for instruction virtualization is de-privileging; for memory virtualization, is the shadow page table. To keep the hypervisor simple and efficient, we adapted pre-virtualization, which achieves the same performance as para-virtualization, at a little amount of engineering cost [4]. By introducing the single address space design, we will argue that address space compression problem can be avoided, VM exit/entry/switching cost can be minimized, and hypervisor protection is achieved.

2.1 Address Space

Figure 1(a) shows KVM's address space layout in which the VM has independent space different from the hypervisor space. KVM hypervisor shares a single address



(a) KVM (b) Single address space hypervisor

Fig. 1. Address space layout for hypervisors (a) KVM (b) Single address space hypervisor

space with VM controller, where VM controller resides in the user space, and the hypervisor in the kernel space. In this design, every VM-exit/-enter causes address space switching between VM and the hypervisor.

The VM controller is an Android(or Linux) task creating a VM, loading guest OS images inside the VM, patching the sensitive instructions of the guest OS, and starting/stopping/pausing/migrating the guest OS. The hypervisor is a kernel module trapping and emulating the sensitive instructions during VM execution.

Figure 1(b) represents our address space layout that enables to design and implement efficient pure software hypervisor. The unique feature in our design is that the VM address space is restricted to the first 4 GB range (0 ~ 4 GB), the first part of huge 64-bit address space. The VM controller is placed anywhere between above 4G and below hypervisor space and the hypervisor resides in the upper part of the 64-bit address space as KVM does. Since we restricted VM address space to 32-bit space, guest OS must be built for 32-bit ARM architecture target. However, the hypervisor and the VM controller should be 64-bit programs.

This address space layout illustrated in Fig. 1(b) looks similar to that of Xen, which is para-virtualization technology, placing Xen hypervisor at the upper part of the VM address space. The Xen guest OS should understand hypervisor resides somewhere within its virtual address space and modify OS code to reserve the upper space to the hypervisor. Our design is different from Xen's in that it does not require such a modification without scarifying performance.

2.2 Address Space Compression

We adopted pure software hypervisor design that has destined to suffer from address space compression problem if we just follow the conventional design. In that design, the hypervisor must use some amount of guest OS's virtual address space which stores VM context, hypervisor context, address space switching functions, and interrupt vector table in order to control the guest OS [3]. That *control space* is mapped into both the guest OS's address space and the hypervisor's, but it is not visible to the guest OS. KVM/ARM calls the control space as the shared page [2]. The address space compression problem is that the guest OS is not allowed to use the shared page space [3].

Under our design assumption, we have extra address space where guest OSs cannot access as well as recognize intentionally or accidentally, but the processor can do. We can avoid compressing the address space of the guest OS by using that extra address space, which implies that we don't need to penetrate into the guest OS's virtual address space and steal some amount of the address space. We place the control space outside of the guest OS's address space, and call it as *gateway* because the control goes to and comes from the guest OS though the gateway. Even though the gateway is not inside the guest OS, it can intercept exceptions occurring inside the guest OS and give a control to the guest OS.

2.3 Protection

The hypervisor including the VM controller must be protected from the guest OSs to keep an entire system safe. Our design presented in Fig. 1(b) ensures that it does not

require any software mechanism to protect the hypervisor from the guest OS. The processor does not allow for the guest OS running in 32-bit compatibility mode to access to the gateway living in higher address than 4 GB because the processor overrides upper 32-bits of the address the guest OS generates. Thus, VM controller and hypervisor living in above 4 GB address can be protected from the guest OS by hardware.

Our design achieves VM isolation, protecting one VM from another VM, by switching only the 32-bit address space. It is like switching the address space of OS tasks to isolate one task from another. The hypervisor only needs to update VM address space of the shadow page table whose hypervisor address space is shared with all the VMs.

While the design confines the guest OS to the 32-bit address space, the processor can access the entire 64-bit address space including interrupt vector table address. When an exception occurs during VM execution, the processor changes its mode to 64-bit mode and moves control to the interrupt vector table entry residing outside of the guest OS in order to handle that exception, by hardware. If the control transfers to the hypervisor, it is allowed to access the VM address space without extra cost because the VM rents lower 32-bit address space that the hypervisor creates and manages.

2.4 VM Exit/Entry/Switch

In our design, no address space change is required but the hypervisor should save and restore VM context and its own context before and after handling the VM exit, and during VM switch since the hypervisor and the guest OS share a single address space. We can reduce the context switching cost due to the 32-bit restriction to the guest OS. That cost includes TLB invalidation, cache invalidation, address space identifier tracking as well as VM context save and restore. Due to 32-bit restriction and the single address space layout, the cost can be minimized and removed respectively.

Our design restricts the execution of the guest OSs to the user-level 32-bit compatibility mode, those OSs and their applications running inside the VM are forced to use user mode AArch32 registers which are subset of AArch64 registers [5]. Even though AArch64 provides 31 general purpose registers(x0 ~ x30), the VM is allowed to use only 16 registers(x0 ~ x15). This restriction makes VM context save/restore cost low. The hypervisor needs to save and restore only w0 ~ w15 general purpose registers for the VM exit and entry respectively. In addition to that benefit, the hypervisor context is not to be saved or restored when switching between the guest OS and the hypervisor because the hypervisor requires that only x19 ~ x29 registers be preserved across the guest OS execution and the guest OS is not allowed to use those registers at hardware level.

3 Prototype Implementation

3.1 Implementation on Nexus 6P

We make the best use of Android task's address space to implement the single address space. We divide 64-bit virtual address space into three parts: low-user(0∼4 GB), high-user(4 GB∼512 GB), and kernel(upper 512 GB). We mapped the low-user, the high-user, and the kernel part to the 32-bit guest OS, the VM controller, and the hypervisor, respectively. The gateway will be installed in the kernel part where the hypervisor is responsible for managing.

As mentioned earlier, the portable hypervisor consists of two modules: user-level task (VM controller) and kernel module (Hypervisor). The controller creates a single address space, evacuates the first 4 GB of the address space and fully assigns to a VM. When a user wants to run a guest OS, the controller loads that guest OS's image into VM address space and patches the sensitive instructions. The hypervisor installs the gateway to catch exceptions during VM execution before the original Android kernel catches. To do that, the hypervisor module set VBAR(Vector Base Address Register) to our own vector table in the gateway area before entering VM. When the hypervisor catches an exception, it analyses the cause of the exception and handles the exception accordingly. If the exception belongs to the original Android kernel, the hypervisor delivers it to the Android kernel without any interpretation. Hardware timer interrupt could be that case.

We implemented our design on a Nexus 6P mobile phone. We have done rooting and installed our customized kernel that supports kernel module insertion. We just changed one line of kernel configuration, CONFIG_MODULES = y, for the kernel to support kernel modules. To the best of our knowledge, since most of the kernel supports kernel modules by default, the custom kernel would not be needed to insert our own kernel module.

3.2 Evaluation

We show the proposed hypervisor efficiency by comparing the message passing time between Linux processes through Linux domain socket and the time between Linux process and echo task over uCOS-II as 32-bit guest OS (Fig. 2). All experiments were done on a Nexus 6P(snapdragon 810 v2.1) and the result is as Fig. 3.

As shown in Fig. 3, in spite of the overhead of proposed hypervisor and uCOS-II, the message passing time is shorter in case that the message size is smaller than 1 K. It takes only 500 processor cycles for 1-byte message, which implies that virtualization cost will be minimized and our design has potential to achieve good performance competitive to KVM.

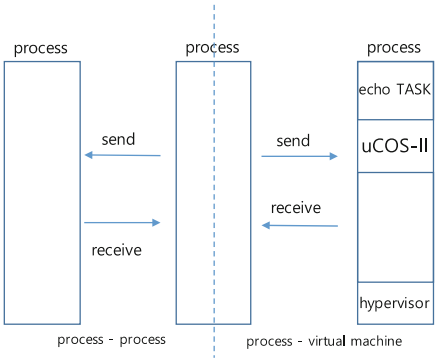


Fig. 2. Experiment environment

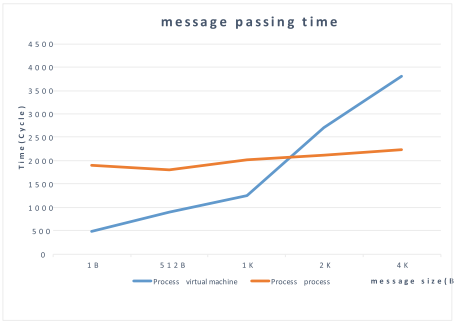


Fig. 3. Comparison of message passing time

4 Conclusion

We introduced the hypervisor design that has portability to any 64-bit Android devices while showing good performance. The main idea in the design is to confine the guest OS to the 32-bit address space. With this idea, we can avoid the address space problem, minimize major virtualization costs (VM exit/entry/switch cost), and achieve to protect the hypervisor from the guest OS.

We believe that the independent 32-bit address space (VM space) is enough to contain core data and service routines that need to be hidden from the main OS. We recommend that most of the service data and routines be placed in the tasks of the main OS and communicate with core routines inside the VM to execute service logics. The single address space design can be used for Intel and AMD processors as well as ARM processors.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0101-16-0644, Research on High Performance and Scalable Manycore Operating System).

References

1. Dall, C., Nieh, J.: KVM/ARM: the design and implementation of the linux ARM hypervisor. In: 19-th International Conference on Architectural Support for Programming Languages and Operating Systems, pp. 333–347 (2014)
2. Dall, C., Nieh, J.: KVM for ARM. In: 12-th Annual Ottawa Linux Symposium, Ottawa, pp. 45–56 (2010)
3. Uhlig, R., Neiger, G., Rodgers, D., Santoni, A., Martins, F., Anderson, A., Bennett, S., Kägi, A., Leung, F., Smith, L.: Intel virtualization technology. IEEE Comput. Soc. **38**(5), 48–56 (2005)
4. LeVasseur, J., Uhlig, V., Yang, Y., Chapman, M., Chubb, P., Leslie, B., Heiser, G.: Per-virtualization: software layering for virtual machines. In: Computer Systems Architecture Conference, pp. 1–9. IEEE (2008)
5. ARM: ARM Cortex-A Series Programmer's Guide for ARMv8-A version 1.0 (2015)

Concept-Based Compound Keyword Extraction Based on Using Sentential Distance, Conceptual Distance and Production Rules: Calculation of the Keyword Importance

Samuel Sangkon Lee(✉)

Department of Computer Science and Engineering, Jeonju University,
303 Chonjam-Ro, Wansan-Gu, Jeonju, Jonbuk 55069, South Korea
samuel@jj.ac.kr

Abstract. Humans can read a document and conceptually organize its contents into few compound keywords that capture the essence of the topic of a document. Based on this information, this study proposes a method for extracting keywords that gives the gist of a document. It uses a set of academic papers as test data to set up a concept-based production rule for forming compound keywords even when author-provided keywords do not appear in the text body of a document. It also proposes a method of calculating the importance of keyword in order to refrain from extracting meaningless keywords. Also the validity of extracted keywords was tested using a data set of thesis paper titles and summaries in the field of natural language processing and speech recognition. Comparison of the author-provided keywords to the keyword results of the developed system showed that the developed system was very useful with an accuracy rate as good as up to 96 %.

Keywords: Keyword pattern · The relation of sentential distance and conceptual distance · Concept word with co-occurrence

1 Introduction

Previous studies on keyword extraction were based on an assumption that the words that capture the essence of a document always appear in the document [2]. The extracted key terms were ranked in order of importance. However, this is not effective when there exists no terms within the document that can be used as a valid keyword, when the potential keywords appear in too many places in the document, or when the keywords are made of too abstract terms that cannot be associated with the subject of the document (the word "extraction" was used in [3] as in terms of derivation, and implication).

For this problem, Nagata et al. defined an indexing rule that states key concepts and co-occurrence of keywords, and proposed a method to generate subject keywords using the defined rule [1]. First step was to extract only nouns from a sentence, and next was to gather synonyms and related terms of such nouns to extract as appropriate keywords. But since the conceptual correlations between the keywords were not considered during

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_69

the extraction of key concepts, the keyword may always not accurately reflect upon the subject of the document.

Therefore this study attempts to develop new method of keyword extraction that uses few key terms arising out of the document while considering their conceptual correlation. In case where no appropriate terms are available within the document, it proposes a concept-based method that can extract compound keywords to describe the subject instead. The compound keywords can be generated using the production rule even when author-provided keywords do not align with field-associated terms [3]. The calculation of the keyword importance is also proposed by focusing on the conceptual correlation of the production rule.

2 Keyword Pattern

Let us look at human-provided keywords patterns as characterized in Table 1 in order to extract keywords that precisely describe the document using such patterns. This table shows 6 examples of sentences or word orders along with extraction patterns. In order to extract appropriate keywords, keyword patterns can be categorized into following three cases. First case is when the keywords all appear within the document, second case is when only some of the keywords appear within the document, and last is when

Table 1. Keyword patterns

Case no.	Sentence example	Extracted pattern	Note
(1)	Speak language and recognize it	Language recognition	Extraction via demonstrative noun
(2)	It aims to process human voice with a processor. First it needs to properly recognize the voice	Speech recognition	Extraction using words appearing in multiple sentences
(3)	Word mining	Keyword extraction	Extraction via compound keyword transformation (Use of thesaurus)
(4)	Man wants to process its language through a machine. They have been working for many decades in order to properly recognize their own language	Speech recognition	Extraction via co-occurrence in multiple sentences
(5)	Inferred knowledge	Artificial intelligence	Extraction via abstract keywords or induced theme
	Can attribute part of speech	Morphological analysis	
(6)	back-off	back-off	Extraction from abbreviations or coined terms
	Context-Free Grammar	CFG	

none of the keywords appear in the document. The six extraction patterns can be precisely formed if the concept of each keyword is utilized. For rule-based keyword production, the compound keywords are dismembered into each keyword for pattern analysis. Since examples of Table 1 require use of transfer dictionary it is left out from the scope of this study. Next chapter proposes a method of keyword extraction using the concept of each word.

3 Keyword Extraction

3.1 Production Rules

Compound keywords are often not directly found in the document and need to be inferred or speculated from words that directly appear in the document. This section defines the relationship between the concept arising from morpheme composite of the keyword and the keyword as a rule based on Nagata's method [1] in attempt to extract abstract or theme-defining words.

When there is a compound keyword w and you assume that morpheme composite of the compound is formed with sub-conceptual words of w, these morpheme composites are concept elements of the word w. These concept elements are made of synonyms and related terms of w. The compound keyword w can be rearranged by grouping synonyms and related terms of each element. The grouping can be structured using brackets as followed where synonyms are contained in the first set of brace and related terms are contained in the second set of brace. Each composite element of w can be divided as $w_1, w_2, \ldots w_n$ and the function Concept() that gets the concept of w can be described as a concept element of each morpheme.

- Concept(meaning) = [{definition, value,…}, {context,…}]
- Concept(processing) = [{dispensing, disposing,…}, {solving, handling,…}]
- Concept(vocal) = [{voice,…}, {timbre, intonation,…}]
- Concept(dialogue) = [{conversation, discussion,…}, {talk,chat,…}]

The production rule (PR) of $w_1, w_2, \ldots w_n$ is to be defined as the summation of concept elements, PR($w_1, w_2, \ldots w_n$) = Concept(w_1) + Concept(w_2) + … + Concept (w_n). This extracts the compound keyword w, or set of $w_1, w_2, \ldots, w_n$ only when all concept elements from Concept(w_1) to Concept(w_n) appear in the document. For example a compound keyword "meaning processing" is only produced if both concepts of "meaning" and "processing" appear in the document, and "spoken dialogue processing" is only produced if all concepts of "spoken", "dialogue", and "process" occur in the document as followed.

- PR(meaning processing) = Concept(meaning) + Concept(processing)
- PR(spoken dialogue process) = Concept(spoken) + Concept(dialogue) + Concept (processing)

The compound keywords formed by the production rule are called the keyword candidates. The production rule is used to restrict generation of keyword candidates unrelated to the context of the document.

3.2 Importance of Keywords

In order to enhance the accuracy of the extraction, a new method of calculation of keyword importance is necessary. To calculate the keyword importance you need a sentential distance, *sd* and conceptual distance, *cd*.

3.3 Conceptual Distance

The sentential distance of sentences that contain each concept element can be used as the conceptual distance, but this distance alone cannot determine the relations between the concept words. Therefore, the conceptual distance is obtained using the co-occurrence of concept words. For example, let us take a look at PR(*XZ*) = Concept (*X*) + Concept(*Z*). The sentential distance, *sd* is as marked by arrows in Fig. 1. Variables v, x, y, z are surface level words that appear in the document and they represent concept elements. Concept words are marked by capitalized alphabets V, X, Y, and Z. The conceptual distance between X and Z is simply the sentential distance 5. The sentential distance, *sd* from sentence *i* to *j* is defined by *s(j) − s(i) + 1* (where $j \geq i$).

Therefore, the *sd* of marked by dashed line in Fig. 1 is 5 (= 5−1 + 1). But since there is a co-occurring concept word *V* between *X* and *Z*, the conceptual distance *cd* becomes 1. The cd is calculated using the following equation.

$$cd = \frac{(cd_1 + cd_2 + \cdots + cd_n)}{cc}. \tag{1}$$

Here, *cc* is the number of common conceptual word *X*, *cd* of *XV* is 1 and *cd* of *VZ* is 1. Therefore, the *cd* becomes $2((= \frac{1+1}{1}, XV, \text{and } VZ)$. This calculates the *cd* of *XZ* while capturing the relation of concept meanings.

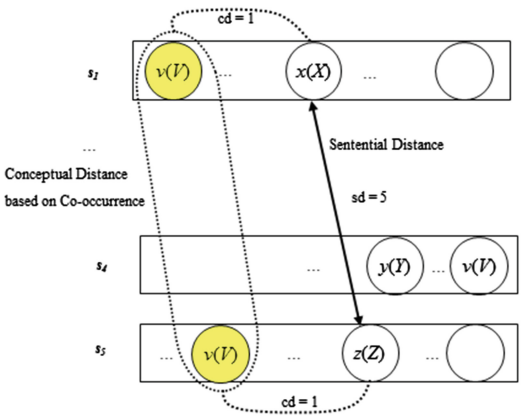


Fig. 1. Relation of sentential and conceptual distance

3.4 Number of Co-occurrences

Concept words or concept elements that represent the main topic usually appears with high frequency within the document. Such concept words are related by co-occurrence with other words. Therefore, it is very important to use a concept word that has a large number of co-occurrences with other words to capture the main topic of the document. Figure 2 describes the co-occurrences in the document where concept words with co-occurrences are connected by a dashed line. If in some document the *i-th* complex word *w* has number of co-occurrence *N(wi)*, the number of co-occurrence of concept word *V*, and *N(V)* becomes 3(= 1 + 1 + 1) as in Fig. 3. Also since *V* is the only concept word that is related by co-occurrence with *X, Y*, and *Z*, *N(X), N(Y),* and *N(Z)* all become 1. Therefore, *N(V)* > *N(X), N(Y)*, or *N(Z)* and *V* has higher importance than that of the *X, Y,* and *Z.*

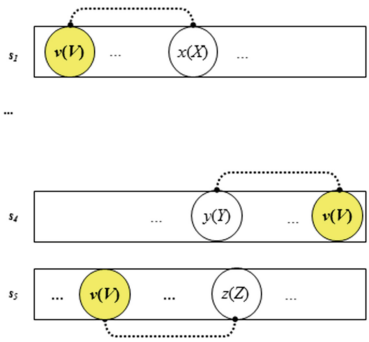


Fig. 2. Conceptual word with co-occurrence

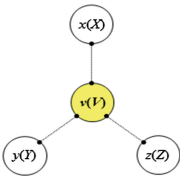


Fig. 3. Number conceptual word with co-occurrence

3.5 Importance Calculation

Calculation of the importance of keyword candidates that consider the conceptual distance *cd* and number of co-occurrence of concept word, *N(wi)* can be expressed as in Eq. (2). The smaller the *cd*, the larger the number of co-occurrence, and the higher the frequency of synonyms and related terms of concept elements, the greater the importance, *I*.

$$I = \left[\frac{1}{n \times cd}\right] \times \sum_{i=1}^{n} \left[\langle \frac{\{(S(w_i) \times \alpha) + (R(w_i) \times \beta)\}}{(S_T \times \alpha) + (R_T \times \beta)} \rangle \times N(w_i)\right]. \quad (2)$$

Here, *n* is the number of concept words that make up the keyword candidates, *S(wi)* is the frequency of synonyms of *wi*, *R(wi)* is the frequency of related terms of *wi*, *ST* is the overall frequency of the synonyms, *RT* is the overall frequency of the related terms, α and β are weights to synonyms and related terms respectively (where, $\alpha > \beta$).

4 Experimental Result and Evaluation

The test data used in this study was a 5,400 set of titles and summaries of Master's and PhD thesis papers (about 80.5 MB) on natural language processing and speech processing selected from a database provided by KTSET (Korean Test Set) and University Library. Table 1 summarizes results of both the keywords extracted by the system and not extracted by the system. Author provided keywords that did not appear in either the title or the summary were treated as answer keywords and were extracted. The test results are as followed. Of 2,010 (average number of four author keywords in one set of title and summary) author-provided keywords there were 1,540 answer keywords. About 70 % of total keywords were compound keywords, and it had an average of two composite morphemes. This shows that the humans often use compound keywords to capture the topic of a document. There were 114 keywords extracted using the production rule and 500 keywords generated using the dictionary (6 in case of Table 1), so total are 2,190. 3,350 keywords were not extracted. This means that 17.5 % of the papers did not have enough trails to generate keywords, and 47.5 % lacked the concept elements (synonyms, related terms) in the dictionary. Production rule was not available at all for the 35 %.

Of the printed keywords, the achieved accuracy rate were up to 96 % for top seven keyword candidates, which is more superior when compared to the results of Nagata. The production rule used in the experiment was tested against 30 basic field terms and subject language. For the dictionary of concept elements, a Korean translated version of classification dictionary put together by Katokawa was used. For those technical terms in the field of natural language processing and speech recognition not listed in the classification dictionary, most similar terms were handpicked from the dictionary. These synonyms and related terms were used as the concept element of that word. The weighting of synonyms and related terms during the experiment were respectively $\alpha = 1$ and $\beta = 0.5$, which were the most optimum number determined from pre-experiments and heuristic algorithm (Fig. 4).

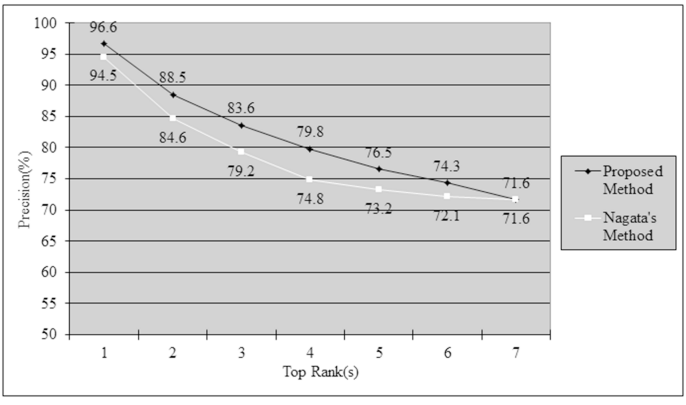


Fig. 4. Comparison of proposed method and Nagata's method

5 Conclusion

Our system scans a document and quickly returns keywords with highest values of importance, so the user can promptly decide if the document matches his/her search query. It also proposes a new method of concept-based compound keyword extraction by focusing on the field-associated terms [3] or keywords that implicate the subject area of the document [4]. In order to improve the accuracy rate of extracted keywords, it defines a concept-based production rule, and proposes a method of calculation of keyword importance using frequency count, co-occurrence, and conceptual distance. The biggest advantage of the developed method is that it can extract not only the author-provided keywords, but also keywords that do not appear in the text body or database [6] of document.

References

1. Nagata, M. et al.: A newspaper keyword gene- ration method based on key-concept extraction. In: Proceedings of the 37th National Convention Information Processing, pp. 1030–1031 (1988)
2. Morohashi, M.: Automatic indexing survey. Mag. IPS Jpn **25**, 918–925 (1984)
3. Lee, S.S., Shishibori, M., Sumitomo, T., Aoe, J.-I.: Extraction of field-coherent passages. Intl. J. Inf. Process. Manage. **38**, 173–207 (2002)
4. Al-Hashemi, R.: Text summarization extraction system (TSES) using extracted keywords. Intl. Arab J. e-Technol. **1**, 164–168 (2010)
5. Chena, Y.-H., Lub, E.J.-L., Tsaib, M.F.: Finding keywords in blogs: efficient keyword extraction in blog mining via user behaviors. Expert Syst. Appl. **41**, 663–670 (2014)
6. Choi, Y.-L., Jeon, W.-S., Yoon, S.-H.: Improving database system performance by applying NoSQL. J. Inf. Process. Syst. (JIPS) **10**, 355–364 (2014)

Coarse-Grained 2.5-D CSAMT Parallel Inversion Method Based on Multi-core CPU

Lili He[1], Hongtao Bai[2(✉)], Jin Wang[3], Yu Jiang[1], and Tonglin Li[4]

[1] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[2] Center for Computer Fundamental Education, Jilin University, Changchun 130012, China
baihongtao@263.net
[3] Information Engineering College, Yangzhou University, Yangzhou 225009, China
[4] College of Earth Survey Science and Technology, Jilin University, Changchun 130026, China

Abstract. In this paper, a 2.5-D controlled source audio-frequency magnetotelluric (CSAMT) reversed parallel method based on multi-core CPU is achieved, aiming to the low efficiency of the large-scale computation in geophysical exploration. This parallel algorithm allocates the different waves to each thread of cores under the coarse-gained mode. The master thread deal with all parallel tasks and other slave threads compute the electromagnetic field values of each wave in parallel Fork-join model. The experiments demonstrate that our parallel algorithm can not only acquire the effective data accuracy, but also obtain about three times than the serial version.

Keywords: 2.5-D inversion · EM field decomposition · OpenMP · Coarse-grained parallelization

1 Introduction

Controlled Source Audio-frequency Magnetotelluric (CASMT) [1–3] uses some powerful artificial sources successfully to explore some fields, such as metal ore, oil, the resources of geothermal, groundwater, etc. The advantage of CASMT is that it is not easily affected by the industries, and more importantly, it has high reliability. Recently, with the development of geophysical exploration in depth and more and more complicated data processing, researchers also draw higher demand in the accuracy and efficiency of numerical simulation. According to some engineering research [4, 5], the computing efficiency of high-dimensional forward and inversion, especially inversion mainly constraints the development of CSAMT.

Coggon [6] firstly proposes a finite element method to solve 2.5-D electromagnetic problem and then the combination of the secondary field iterative solution and the infinite elements has considerably reduced the computational consume in [7]. Stoyer and Greenfield [8] apply a different finite method to compute an oscillating magnetic dipole source, which has already employed successfully in exploring the ground water

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_70

resources. In [9], to image the resistivity of the subsurface, the interested area is divided into some grids by using the Greens function, Mao and Bao deduce a Boundary Integral Equation in which the voltage on the subsurface satisfies the condition of arbitrary conductivity distribution in 2.5-Dimensional problem. Di and Unsworth [10] adopts 2.5-D forward modeling and uses sensitivity matrix calculated approximately by CSAMT-RRI method to solve 2.5-D finite element CSAMT numerical inversion. When analyzing the complex geoelectric model, 2.5-D finite element numerical simulation need disperse and divide many structural units, and repeatedly solve equations during the wave domain, which brings a huge challenge to the memory and computing of computers. How to apply high-performance computers and parallel analysis techniques to improve the effectiveness of high dimensional electromagnetic numerical simulation is a key step. The MPI method is used to parallel rapid relaxation inversion of 3D magnetotelluric data [11]. The staggered grid finite difference method on parallel computation is utilized for 3D magnetotelluric modeling in [12]. The parallel 3D magnetotelluric forward modeling algorithm is introduced in [13]. However, there are few algorithms for CSAMT, especially the parallel inversion algorithm.

In this paper, we analyze the disadvantage of previous 2.5-D CSAMT inversion algorithms and then assign multiple wave inversion computation tasks to some cores of the modern CPU. The master thread is responsible for assigning tasks to some other threads whose task is inversion computation. The master thread can share the memory and exchange the data with these other threads. We test some different sizes of geo-electric models and show the relationship between the effectiveness of parallel and the number of cores and threads. Experimental results show that our method can acquire the same quality of solution, compared with the sequence methods. Moreover, our method in 4-cores CPU gains 3 times speed-up over the sequence methods.

2 The Inversion Theory of 2.5-D CSAMT

Inversion uses the results of forward modeling to simulate the observed values of surface, and further describe the electrical structure of underground medium. Forward model applies Cole-Cole [14] to describe the complex resistivity of each small unit in the underground network subdivision, which uses the damped least square method for linear inversion.

f_{si} is denoted as the value of observation field, $F_i(r)$ is denoted as the value of forward field, $\overline{X}$ is denoted as the four parameters of Cole-Cole model of underground medium units (zero frequency resistivity ρ_0, polarization m, time constant τ, the correlation coefficient of frequency c), and $\delta_i(\overline{X})$ is denoted as the relative deviation of fitting between theory and the measured values.

$$\delta_i(\overline{X}) = [f_{si} - F_i(r)]/f_{si} \tag{1}$$

The inversion fitting error $\Phi(\overline{X})$ is set as

$$\Phi(\overline{X}) = \sum_{i=1}^{m} [\delta_i(\overline{X})]^2 \tag{2}$$

where $i = 1,\ldots,m$ is the position or frequency of each observation point.

Since the forward function is non-linear, we do linearization approximation for the deviation function. At first, the model parameters are given a set of initial value $\overline{X}^0$, Then $\delta_i(\overline{X})$ in $\overline{X}^0$ uses the Taylor expansion, and at the same time ignores the second and the second above of derivative:

$$\delta_i(\overline{X}) \approx \sum_{i=1}^{m} [\delta_i(\overline{X}^0) + \sum_{k=1}^{n} p_{ik} \cdot \Delta x_k]^2 \tag{3}$$

The fitting error is set as the multivariate function Δx_1, Δx_2,…, Δx_n of the modifier of conductivity model and its minimum conditions is that:

$$\frac{\partial \Phi(\overline{X})}{\Delta x_j} = 2 \sum_{i=1}^{m} [\delta_i(\overline{X}^0) + \sum_{k=1}^{n} p_{ik} \cdot \Delta x_k] \cdot p_{ij} = 0 \tag{4}$$

Further derivation:

$$\sum_{i=1}^{m} \sum_{k=1}^{n} p_{ij} p_{ik} \Delta x_k = -\sum_{i=1}^{m} \delta_i(\overline{X}^0) \cdot p_{ij} \tag{5}$$

where $j = 1,2,\ldots,n$, the modifier of linear equations can be deduced, as shown in the formula 6:

$$(P^T P + \lambda I) \cdot \Delta X = S \tag{6}$$

P is Jacobi matrix, whose element is p_{ik}, $\Delta X = (\Delta x_1, \Delta x_2,\ldots, \Delta x_n)$, I is denoted as an unit matrix, the right-end vector $S = (s_1,s_2,\ldots,s_n)$, $s_j = -\Sigma_{i=1}^{m} \delta_i(\overline{X}) p_{ij}$, λ is set as the constant damping factor.

3 Parallel Algorithm

3.1 Serial 2.5-D CSAMT Inversion Algorithm

According to the moving source, frequency, the number of wave in each loop, 2.5-D CSAMT sequence algorithms compute linear equations (formula 6) respectively based on the different receivers. Therefore, if the solved formula 6 can be parallized, this algorithm can improve the efficiency of serial algorithms. The previous analysis finds that, the main reason of solving linear equations under costing a lot of time by the inversion algorithm is that the large number of iteration, rather than a single long iteration time, thus a single iteration parallel algorithm as the fine-grained parallel is not

suitable for 2.5-D CSAMT method. In the outer layer of the formula (6), in the loop of wave number, frequency, and mobile source, the number of wave is set to from 15 to 20, and both of the number of frequency and mobile source is from 3 to 8. Furthermore, the calculation of wave, frequency and mobile source is completely independent. Therefore, we can parallel the wave, frequency and mobile source, which is feasible.

3.2 The Parallelization Based on OpenMP

With the development of high-performance computing, how to fully play to the efficiency of modern microprocessor or processor is a key problem. It is hard to improve the integration level, control volume and power consumption. The main way to follow the Moore's law is to develop a CPU core within more and more cores and supporting the hyper-threading transformation. The current commercial mainstream microprocessor has already supported for eight or more processor cores. For this consideration above, we make use of the advantage of shared memories, and apply OpenMP in multi-core of CPU for the computation of hyper-threading. Considering that the number of wave is relatively fixed and large, it is advantageous to develop the performance of multi-core CPU dynamic scheduling threads.

In this paper, we use a wave coarse-grained parallel method, and this parallel 2.5-D CSAMT inversion algorithm is proposed. In detail, this algorithm uses Fork-join model. The main process reads the parameters of model. After generating subdivision relationship between the nodes, the guidance statements are compiled by OpenMP. The number of waves compute by dynamically allocated to each thread, and then each independent thread computers the linear equations of coefficient matrix, right-end and linear equations. In the process of computing auxiliary field and sensitivity, each thread shares the memory regions of the auxiliary field and sensitivity matrix, which will be written back to the results. After computing the number of wave, threads will be ended by the compile statements. At the same time, the main process continues to run the reverse Fourier transform and the error fitting, until the convergence of our algorithm.

4 Experimental Results

In this section, we select many inversion model based on different types and scales. We test these models, in terms of both of the computational accuracy and the computational efficiency. Our experiment uses Intel Ivy Bridge Core i5 3.2 GHZ with 4 G memory.

4.1 The Computational Accuracy

Firstly, we generate the abnormal body model, where the network subdivision is set to 55×28, the layer of air subdivision is set to 10, the layer of land subdivision is set to 18, and the number of wave number is set to 18.

Let the value of the air resistivity is 10^6 Ωm and the value of earth resistivity is 100 Ωm. We use six mobile sources, where the centers of sources are located in the 13, 15, 17, 19, 21, and 23 subdivision of earth surface respectively. They place along the

direction of the source, and both of the lengths of the emission line source and receiving a dipole are 100. The number of frequency is 6, including 0.02 Hz, 0.1 Hz, 1.0 Hz, 8.0 Hz, 32.0 Hz, 128.0 Hz. The number of observation point is 10, whose location is at the surface (z is set to 0), x is set to from 200 m to 1100 m. The distribution of abnormal body: the distribution range of x shaft is from 500 m to 700 m, the distribution range of z shaft is from 100 m to 400 m. The initial values of the complex resistivity model Cole-Cole: the zero frequency resistivity is 100.0 Ωm, the frequency coefficient is 0.3, the time constant is 10.0 s, and the polarization rate is 0.3. We present the comparison of serial and parallel inversion shown in Fig. 1.

As you can see from Fig. 1, the results of parallel and serial inversion methods are same, which proves this proposed parallel method is reliable and effective.

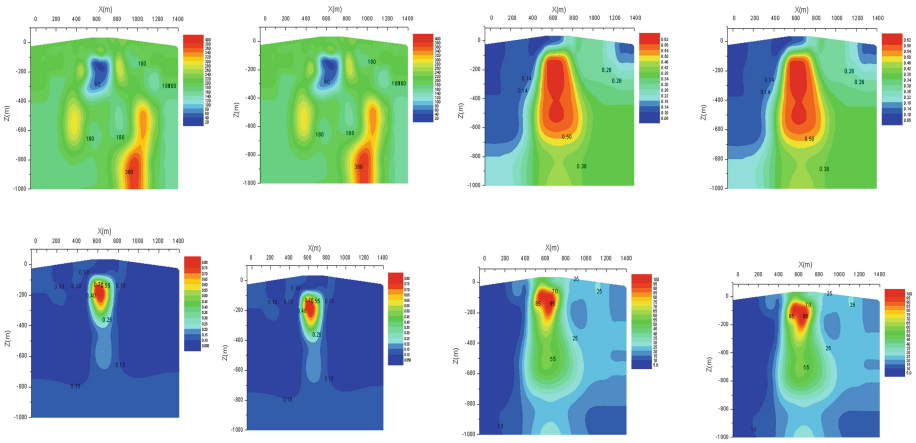


Fig. 1. The Inversion results of zero frequency resistivity, frequency coefficient, polarization, and time constant in 20th iteration (left is serial and right is parallel)

4.2 The Speed-Up Performance

We test the performance of the parallel and sequence methods on the single abnormal body model. Specially, the model range is from 100 × 14 to 100 × 56. The number of mobile sources, observation points and frequency is 2. The fixed number of iterations is set to 10. The number of threads is set to from 1 to 8, where the super threads are from 5 to 8.

Table 1 shows that the performance of our method is improved followed by increasing the number of threads. When the number of thread reaches 7, the improvement is best. However, when the number is 8, the performance is not good. This is because when the number of wave is 18, including the eight hyper-threading threads, the time consumption of synchronization offsets the gain performance.

Table 1. Computation results of single abnormal body model

Network subdivision	|Thread|	Run time (s)	Speed-up (Sp)	The parallel efficiency (Ep)
100 × 14	1	800	Null	Null
	2	553	1.45	72.33 %
	3	407	1.97	65.52 %
	4	349	2.29	57.31 %
	5	307	2.61	52.12 %
	6	265	3.02	50.31 %
	7	261	3.07	43.79 %
	8	283	2.83	35.34 %
100 × 28	1	2187	Null	Null
	2	1569	1.39	69.69 %
	3	1142	1.92	63.84 %
	4	962	2.27	56.83 %
	5	845	2.59	51.76 %
	6	733	2.98	49.73 %
	7	718	3.05	43.51 %
	8	776	2.82	35.23 %
100 × 56	1	6336	Null	Null
	2	4627	1.37	68.47 %
	3	3366	1.88	62.75 %
	4	2941	2.15	53.86 %
	5	2528	2.51	50.13 %
	6	2178	2.91	48.48 %
	7	2114	3.00	42.82 %
	8	2292	2.76	34.55 %

5 Conclusion

In this paper, we use the computing resources of multi-core CPU to design a 2.5-D CSAMT parallel inversion algorithm based on reasonably allocating the parallel granularity. The efficiency of our algorithm outperforms the serial algorithms under the same data accuracy. Moreover, we design the coarse-grained parallel based on the number of wave. According to the characteristics of our algorithm, the mobile source and frequency are completely independent. By using a computer cluster technology, we try to parallel nodes between in moving source and frequency domain, or wave domain, which will further enhance the efficiency of our inversion algorithm. In addition, many algorithms based on GPU increasingly play an important role in [15]. It may be a good way to research the geophysical prospecting GPU parallel computing method.

Acknowledgements. The authors gratefully acknowledge the financial supports for this research from the National Natural Science Foundation of China (51409117, 61272208, 61572228), Jilin Province Department of Education "Thirteen Five" scientific and technological research projects [2016] No. 432, Fundamental Research Funds for the Central Universities Special (JCKY-QKJC47, JCKY-QKJC49).

References

1. Sutarno, D., Fatrio, I.: Robust M-estimation of CSAMT impedance functions. Indonesian J. Phys. **18**(3) (2015)
2. Wang, R., Yin, C., Wang, M., et al.: Laterally constrained inversion for CSAMT data interpretation. J. Appl. Geophys. **121**, 63–70 (2015)
3. Da, L., Xiaoping, W., Qingyun, D., et al.: Modeling and analysis of CSAMT field source effect and its characteristics. J. Geophys. Eng. **13**(1), 49 (2016)
4. Yi, H., Xueming, S., Jianke, F., et al.: Review on parallel computing and its application in exploration geophysics. Progress Geophys. **25**(2), 642–649 (2010)
5. Wang, R., Yin, C., Wang, M., et al.: Laterally constrained inversion for CSAMT data interpretation. J. Appl. Geophys. **121**, 63–70 (2015)
6. Coggon, J.H.: Electromagnetic and electrical modeling by the finite-element method. Geophysics **36**(2), 132–155 (1971)
7. Unsworth, J.M., Bryan, J.T., Alan, D.C.: Electromagnetic induction by a finite electric dipole source over a 2-D earth. Geophysics **58**(2), 198–214 (1993)
8. Stoyer, C.H., Greenfield, R.J.: Numerical solutions of the response of a two-dimensional earth to an oscillating magnetic dipole source. Geophysics **41**(3), 519–530 (1976)
9. Mao, X., Bao, G.: A new method for 2.5-dimensional resistivity forward modelling. J. Central South Univ. Technol. **28**(4), 307–310 (1997)
10. Qingyun, D., Miaoyue, W.: 2.5-D finite-element CSAMT numerical inversion. Oil. Geophys. Prospect. **41**(1), 100–106 (2006)
11. Changhong, L., Handong, T., Tuo, T.: Parallel rapid relaxation inversion of 3D magnetotelluric data. Appl. Geophys. **6**(1), 77–83 (2009)
12. Li Yan, H., Xiangyun, Y.W., et al.: A study on parallel computation for 3D magnetotelluric modeling using the staggered-grid finite difference method. Chinese J. Geophys. **55**(12), 4036–4043 (2012)
13. Handong, T., Tuo, T., Changhong, L.: The parallel 3D magnetotelluric forward modeling algorithm. Appl. Geophys. **3**(4), 197–202 (2006)
14. Cole, K.S., Cole, R.H.: Dispersion and absorption in dielectrics I. Alternating current characteristics. J. Chem. Phys. **9**(4), 341–351 (1941)
15. Jermain, C.L., Rowlands, G.E., Buhrman, R.A., et al.: GPU-accelerated micromagnetic simulations using cloud computing. J. Magn. Magn. Mater. **401**, 320–322 (2016)

Virtual Force and Glowworm Swarm Optimization Based Node Deployment Strategy for WSNs

Jin Wang[1], Yiquan Cao[1], Jiayi Cao[1], Huan Ji[1], and Xiaofeng Yu[2(✉)]

[1] School of Information Engineering, Yangzhou University, Yangzhou, China
[2] School of Business, Institute of Electronic Business (Yancheng), Nanjing University, Nanjing, China

Abstract. Wireless Sensor Networks (WSNs) can be viewed as a network with hundreds or thousands of randomly deployed sensors, whose coverage control problem has the characteristics of self-organized groups. This paper studies the coverage optimization strategy based on swarm intelligence for wireless sensor networks. In WSNs, the random deployment of nodes causes the coverage of the blind area and the redundancy of the coverage. We propose a new algorithm based on virtual force and glowworm swarm optimization algorithm. Firstly, the utilization rate of the nodes and the effective coverage of the network are the optimization objectives and the corresponding mathematical model is established. Then, the virtual force algorithm and the glowworm swarm optimization algorithm are used to solve the problem of modeling, and the optimal coverage scheme for WSNs is obtained. Simulation results show that the virtual force and glowworm swarm optimization algorithm can effectively improve the coverage of WSNs nodes, reduce the redundancy of sensor nodes, and reduce the cost of network effectively. Besides, the network survival time can get prolonged.

Keywords: Wireless Sensor Network · Coverage optimization · Glowworm swarm optimization algorithm · Virtual force algorithm

1 Introduction

With the rapid development of information technology, it is very important for people to master the real-time and effective information. Wireless sensor networks (WSNs) [1] are the most popular data collection technology. WSNs have many advantages such as low energy consumption, easy distribution, low cost and self-organization etc. WSNs have been widely used in many applications like real life [2], such as military surveillance, industrial and health care, etc.

The coverage control [3] is a basic problem for various WSNs applications. It is a study of the guaranteeing service quality that how to realize the maximum of the network coverage and provide reliable monitoring and target tracking service under certain conditions. The effective coverage control strategy and the application of the algorithm can optimize the distribution of the various resources of WSNs, which is helpful for the effective utilization of energy in network nodes, the improvement of service quality and the extension of overall survival time.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_71

In recent research, the utilization of swarm intelligence algorithm is becoming more and more popular. Through simulating the behavior of the nature, a kind of stochastic optimization algorithm is established, which is called swarm intelligence algorithm. It is a new intelligent model, produced by the gradual development of computational intelligence in the field of swarm intelligence.

The remaining of this paper is organized as follows. System model is given in Sect. 2. Section 3 presents relevant system mode. We propose our algorithm with detailed explanations in Sect. 4. Experiment results are shown in Sect. 5, and Sect. 6 concludes this paper.

2 Related Work

Many scholars have carried out related research on WSNs coverage issue, among which the most original coverage optimization algorithm is based on the graph theory and detection [4, 5]. But these algorithms have some disadvantages. Graph theory algorithm assumes that a sensor node can be found at any point in the monitoring area, which is not possible for real application. Detection algorithm that only suitable for a certain size of the WSN cannot guarantee the full coverage of the network [6]. More and more scholars apply swarm intelligence algorithm to WSNs coverage optimization.

The authors in [7] proposed that using genetic algorithm to achieve coverage optimization. Genetic algorithm has strong global search ability and has parallel search ability, but the convergence speed of the optimal solution is slow. It is difficult to meet the real-time requirements of the dynamic node selection. The authors in [8, 9] proposed a kind of coverage optimization method based on particle swarm optimization algorithm for WSNs. It is proved that the particle swarm optimization algorithm can achieve the coverage optimization of WSNs effectively. But the standard particle swarm algorithm is easy to fall into premature phenomenon in space search, which limits the searching range of particles.

S. Mini et al. [10] use ABC algorithm to realize coverage optimization, where the selection method of sensitivity and pheromone which replace the original roulette wheel selection strategy complete the search of the neighborhood search. But the convergence speed is slow and easy to fall into local convergence. Authors in [11] proposed OAFSA (Optimized Artificial Fish Swarm Algorithm). The multi-objective optimization problem is shifted to a single objective optimization problem. The algorithm has the defects of low accuracy and falls into local convergence in the later stage.

In view of the above shortcomings, this paper presents a coverage optimization algorithm based on virtual force and the glowworm swarm optimization algorithm for WSNs. Firstly the utilization rate of the nodes and the effective coverage of the network are the optimization objectives and the corresponding mathematical model is established. Then, the virtual force algorithm and the glowworm swarm optimization algorithm are used to solve the model. Finally, the optimal coverage scheme for WSNs is obtained.

3 System Model

3.1 Area Coverage Rate

Assuming the monitored area is a two-dimensional plane, N sensor nodes are arranged in the monitoring area, and each sensor can be covered by a range of r. Sensor node set is $C = \{c_1, c_2, \cdots, c_n\}$, where $c_i = \{x_i, x_j, r\}$ means $\{x_i, x_j\}$ coordinates for the center, the coverage area of a circle radius r. We use the similar Probability perception model as that of [12].

Assuming that the monitoring area is digitally dispersed into $m \times n$ pixels and the area size of each pixel is $\Delta x \times \Delta y$, the coverage rate $p(x, y, c_i)$ indicate whether each pixel is covered by a wireless sensor network node. The monitoring area coverage rate $R(C) = \frac{\int_0^m \int_0^n p(x,y,C)dxdy}{m \times n}$.

3.2 Fitness Function

There are two conditions to be met: the first is to maximize the coverage of the wireless sensor network; the second is to maximize the utilization of nodes in wireless sensor network coverage optimization. So, the fitness functions can be described as follows:

$$F(x) = \omega_1 \times \left(\frac{|c| - |c^i|}{|c|}\right) + \omega_2 \times \left(1 - \frac{\int_0^m \int_0^n p(x, y, C)dxdy}{m \times n}\right), \ \omega_1 + \omega_2 = 1 \quad (1)$$

where ω_1, ω_2 represent the weights of corresponding function which depend on the requirements of the network performance. $|c|$ is total number of deployed sensor nodes and $|c_i|$ is the number of working sensor nodes.

4 The Proposed Algorithm

4.1 Initial Deployment Based on Virtual Force Algorithm

The virtual forces between nodes s_i^k and s_j^f is obtained below:

$$F_{ij} = \begin{cases} (\omega_a(d_{ij} - d_{th}), a_{ij}), & d_{th} \le d_{ij} \le c \\ (\omega_r(\frac{1}{d_{ij}} - \frac{1}{d_{th}}), a_{ij} + \pi), & d_{th} \ge d_{ij} \\ 0, & d_{ij} \ge c \end{cases} \quad (2)$$

where a_{ij} is azimuth angle of node s_i^k and s_j^f, c is distance threshold, ω_a and ω_r represent the gravity coefficient and repulsion coefficient, which is a regulator of node density basis.

In addition to the force of other nodes, the force of the node is also affected by the obstacle and the hot spot area in the virtual force algorithm. Therefore, the node s_i^k updates location from (x_{old}, y_{old}) to (x_{new}, y_{new}) under the virtual force.

$$\begin{cases} x_{new} = x_{old} + \frac{F_{ix}}{F_i} \times Maxstep \times e^{\frac{-1}{F_i}} \\ y_{new} = y_{old} + \frac{F_{iy}}{F_i} \times Maxstep \times e^{\frac{-1}{F_i}} \end{cases} \tag{3}$$

where F_{ix} and F_{iy} represent x-axis and y-axis component of the virtual force F_i, *Maxstep* is maximum moving distance.

4.2 GSO Algorithm

Our proposed VF-GSO (virtual force and glowworm swarm optimization algorithm) used for the deployment of WSNs nodes is as follows:

Step 1: According to the initial deployment of the virtual force algorithm, each sensor node is regarded as a firefly and form the firefly group. Each spilled nodes have the same concentration of fluorescein in the initial algorithm;

Step 2: Update sensor the fluorescein of sensor nodes;

$$l_i(t+1) = (1-\rho)l_i(t) + \gamma F(x_i(t+1)) \tag{4}$$

where $l_i(t)$ represent(s) the fluorescein concentration node i in the t iterations, $\rho(0<\rho<1)$ is fluorescein concentration attenuation coefficient. $F(x_i(t+1))$ is the objective function value of the node i in the t iterations.

$$F(x_i(t+1)) = \sum_{j=1}^{k} \frac{l_j(t+1)}{d_{ij}(t+1)} \quad j \in N_i(t+1) = \left\{ j : d_{ij}(t+1) < r_d^i(t+1) \right\} \tag{5}$$

where d_{ij} is Euclidean distance between i and j, j is a neighboring node in the sensing radius of the node i, r_d^i is the radius of the adjacent decision region in the t iterations.

Step 3: The probability p_{ij} is calculated as follows:

$$p_{ij} = \frac{l_i(t) - l_k(t)}{\sum\limits_{k \in N_i(t)} (l_i(t) - l_k(t))} \quad k \in N_i(t) = \left\{ k : d_{ik}(t) < r_d^i(t),\ l_i(t) < l_k(t) \right\} \tag{6}$$

where k represents node i sensing radius and the concentration of fluorescein is lower than the adjacent nodes of node i in the t iterations, $d_{ik}(t)$ is the Euclidean distance between the firefly i and k in the t iterations.

Step 4: Update the position of sensor node;

$$x_i(t+1) = x_i(t) + s * \left(\frac{x_j(t) - x_i(t)}{\| x_j(t) - x_i(t) \|}\right) \tag{7}$$

where $x_i(t)$ is the location of node i in the m dimensional space, s is step size of location updating, $x_j(t) - x_i(t)$ represents the standard Euclidean distance.

Step 5: Update the radius of the adjacent decision region. If the number of iterations is less than the specified number of times, then go to step 2

$$r_d^i(t+1) = \min\{r_c, \max\{0, r_d^i(t) + \beta(n_t - |N_i(t)|)\}\} \tag{8}$$

where β is a proportional constant, r_c is sensing radius of node, n_t is parameter of the number of neighbors in the control neighborhood range.

5 Performance Evaluation

Assuming that wireless sensor network monitoring area is 50 m * 50 m. There are [5–55] sensor nodes randomly deployed inside the area. The perception distance of each sensor node is [3–8] m, communication distance is 2 times than the perception distance. VF-GSO algorithm parameters are: $Maxstep = 0.5$, $\rho = 0.3$, $\beta = 0.08$, $s = 0.3$, $\gamma = 0.4$, $\omega_1 = 0.3$, $\omega_2 = 0.7$.

In this paper, the VF-GSO algorithm is proposed to solve the problem of sensor node coverage scheme. In order to prove the superiority of the method, the paper compares the results of the methods with genetic algorithm and glowworm swarm optimization algorithm.

The degree of network coverage reflects the integrity of regional monitoring. The relationship between the degree of coverage and the nodes under the three algorithms are shown in Fig. 1. When the number of nodes is small, the coverage of the three

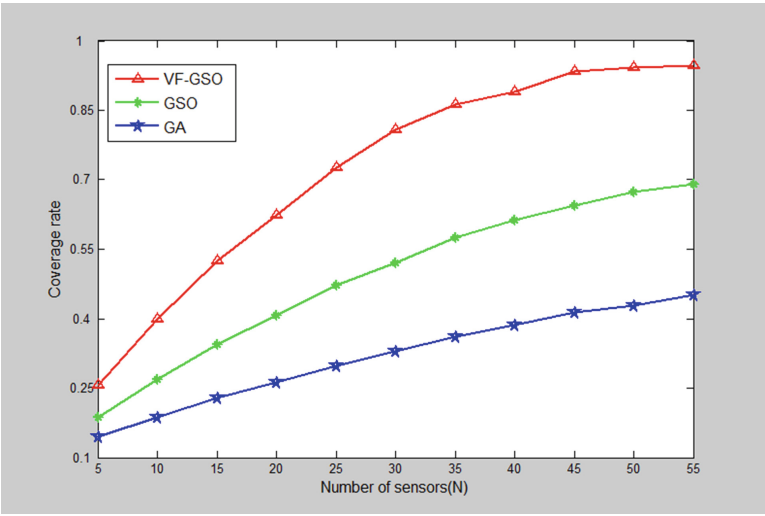


Fig. 1. Comparison of network coverage rate

algorithms is almost the same. Increasing node density, the degree of VF-GSO coverage increases most quickly. The algorithm can reach 96 % and maintain the stability of coverage, when the sensor node reaches a certain level. Compared with the glowworm swarm optimization algorithm and genetic algorithm, this algorithm has been largely improved in terms of network coverage.

The astringency of the network coverage reflects the real-time of regional monitoring. With the increase of the sensing radius, the number of iterations is gradually reduced in Fig. 2. When the sensing radius is 3 m, GA algorithm reached convergence of the iterations is 725, the number of GSO iteration is 480 and the number of VF-GSO is 400. When the radius is 8 m, the number of iterations of the GA algorithm is 384, the GSO is 210 and VF-GSO is 140. We can get the conclusion that the proposed algorithm has better astringency in this paper.

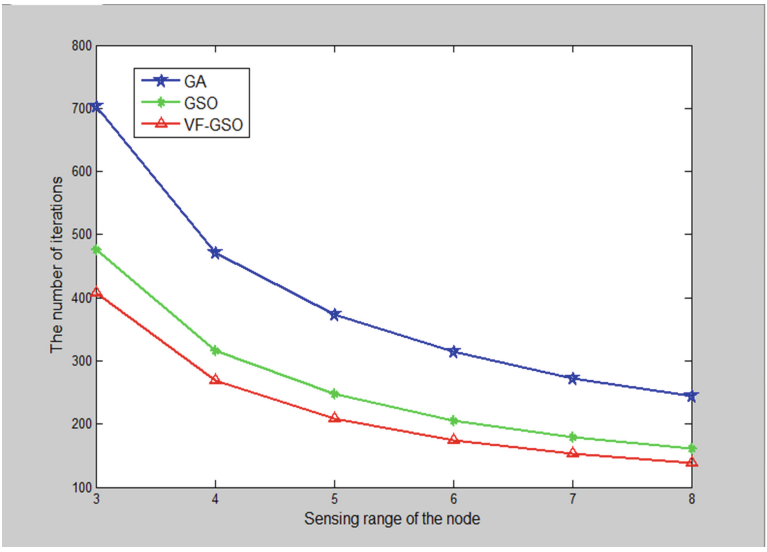


Fig. 2. Comparison of astringency speed

6 Conclusion

We analyze the optimization problem of WSNs coverage and propose a new optimization algorithm based on virtual force and glowworm swarm optimization algorithm, which set up mathematical model of network coverage and node equilibrium. This paper proposes the improvement method of glowworm swarm optimization and the application in the optimization of network coverage. It gives the working steps of algorithm and simulation examples. Simulation results show that the proposed VF-GSO algorithm can achieve better performance than traditional GA and GSO algorithms.

Acknowledgment. This work was supported by the National Natural Science Foundation of China (61402234). Professor Xiaofeng Yu is the corresponding author.

References

1. Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Comput. Netw. Int. J. Comput. Telecommun. Netw. **52**(12), 2292–2330 (2008)
2. Kaur, R., Lal, M.: Wireless sensor networks: a survey 1. Comput. Netw. **38**(4), 393–422 (2002)
3. Wang, X., Han, S., Wu, Y., et al.: Coverage and energy consumption control in mobile heterogeneous wireless sensor networks. IEEE Trans. Autom. Control **58**(4), 975–988 (2013)
4. Wu, C., Si, P., Zhang, Y., et al.: A redeployment strategy based on unmanned aerial vehicle in wireless sensor network. In: 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference (ITAIC), pp. 343–346. IEEE (2011)
5. Onur, E., Ersoy, C., Deli, C.H.: Sensing coverage and breach paths in surveillance wireless sensor networks. Sens. Netw. Oper. **2**, 984–988 (2005)
6. Min, A.W., Shin, K.G.: Robust tracking of small-scale mobile primary user in cognitive radio networks. IEEE Trans. Parallel Distrib. Syst. **24**(4), 778–788 (2013)
7. Hu, X.M., Zhang, J., Yu, Y., et al.: Hybrid genetic algorithm using a forward encoding scheme for lifetime maximization of wireless sensor networks. IEEE Trans. Evol. Comput. **14**(5), 766–781 (2010)
8. Katsenou, A.V., Kondi, L.P., Parsopoulos, K.E.: Motion-related resource allocation in dynamic wireless visual sensor network environments. IEEE Transac. Image Process. Publ. IEEE Signal Process. Soc. **23**(1), 56–68 (2014)
9. Deif, D.S., Gadallah, Y.: Classification of wireless sensor networks deployment techniques. IEEE Commun. Surv. Tutorials **16**(2), 834–855 (2014)
10. Mini, S., Udgata, S.K., Sabat, S.L.: Sensor deployment and scheduling for target coverage problem in wireless sensor networks. Sens. J. IEEE **14**(3), 636–644 (2014)
11. Huang, Y.Y., Ke-Qing, L.I.: Coverage optimization of wireless sensor networks based on artificial fish swarm algorithm. Appl. Res. Comput. **30**(2), 554–556 (2013)
12. Hossain, A., Chakrabarti, S., Biswas, P.K.: Impact of sensing model on wireless sensor network coverage. IET Wireless Sens. Syst. **2**(3), 272–281 (2012)

Comparison with Recommendation Algorithm Based on Random Forest Model

Yu Jiang[1,2], Lili He[2(✉)], Yan Gao[2], Kai Wang[2], and Chengquan Hu[2]

[1] Key Laboratory of Information System Security of Ministry of Education, TNLIST, School of Software, Tsinghua University, Beijing 100084, China
[2] College of Computer Science and Technology, Jilin University, Changchun 130012, China
helili@jlu.edu.cn

Abstract. Product recommendation based on user behavior is a hot research topic In the Internet era in the same data set, the features that the results of the various classifications are a greater difference were handled with random forest model. This paper compares the mainstream classification algorithm C4.5 and CART and analyzes 578,906,480 user behavior records on the results of actual transaction in Alibaba. The results show that CART decision tree algorithm is more suitable for large e-commerce data mining.

Keywords: User behavior · Random forest model · Decision tree · C4.5 · CART

1 Introduction

User implicit demand excavated from the mass of information on user behaviors is essential for service providers. Currently, the recommended system [1] has been preliminarily applied in business, but how to construct a highly efficient and intelligent recommendation algorithm is still a hot topic. Random Forests model that a classification prediction model [2] is proposed by Leo Breiman, it has many advantages, such as learning faster, less parameters and fault tolerance, since it was proposed in many fields received applications. Guo Yingjie et al. used random forest classification to identifies plant resistance gene [3]; Li Jiangeng et al. analyze gene pathways of cancer microarray data based on random forest [4] and Fang Kuangnan predicts fund yields direction used random forests model [5].

In this paper, the dataset is massive amounts of user behavior in the Alibaba website real deal. We defined user behavior attribute set and compared with classification algorithm C4.5 and CART based on random forest model to provide evidence for better user recommendation.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_72

2 Basic Theory

2.1 Random Forests Model

Random Forests is classifier made more decision independent trees [6, 7]. The generation of decision tree is generally controlled by the property division and pruning, but when a large number of features, it may be over-fitting problems. Random forests use boosting [8, 9] resampling method to extract plurality of samples from the original data set, and to construct the decision tree for each sample, through the plural the of decision tree, it can forecast the final prediction results (Fig. 1).

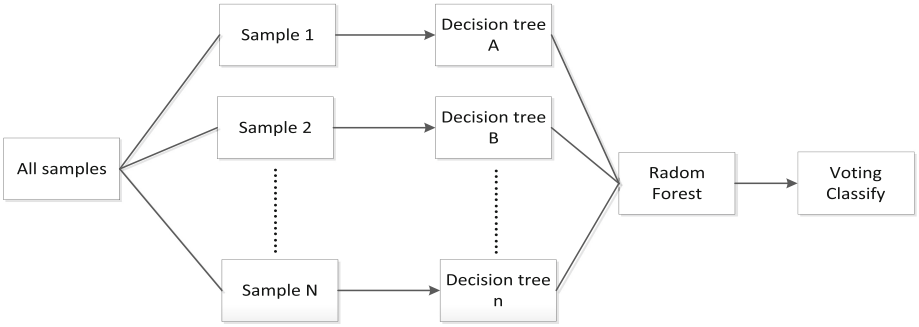


Fig. 1. Random forests model

2.2 C4.5 Algorithm

C4.5 algorithm [10] starting from the root node assigned the best properties. The value of each attribute will generate the corresponding branch, and generate new nodes on each branch. Best attribute selection criteria is based on the definition of information entropy gain ratio to select test properties of the node, entropy characterizes the purity of any sample set. There are four steps to establish C4.5:

(1) Handling the data source to convert the continuous data into discrete;
(2) Calculating its information gain and information gain ratio for each attribute;
(3) The possibility value of each attribute corresponds to a subset, it is start from the root node; Second step is performed recursively until each subset of data gets the same value attributes and generates decision trees.
(4) Extraction of classification rules based on the decision trees can classify new data set.

2.3 CART Algorithm

Classification and Regression Tree (CART) algorithm [11, 12] is a very effective non-parametric classification and regression algorithm, it achieves the purpose of prediction by constructing a binary tree. Binary Tree is not easy to generate data

fragments and its accuracy will be higher than often multi-tree, so we choose binary tree in the CART algorithm. CART with Gini index as the division standard. CART is established by the following three steps:

(1) Creating binary trees used data sets, then disrupting each attribute node until all samples of leaf nodes are classified into the same category or disrupted attribute sets are empty;
(2) Pruning, pruning algorithm is continuously used to get smaller trees and form an ordered sequence of sub-tree;
(3) Selecting the final result, the final decision tree is chosen the best sub-tree from the subset of sub-tree sequence according to the validation.

3 Comparative Analysis

3.1 Experimental Design

Experimental dataset is 578,906,480 recorded data provided by large data competition of Alibaba, The data spans a period of 4 months. the data of each record is by user_id, brand_id, type, visit_time four fields, which user_id uniquely identifies the user; brand_id uniquely identifies brand; type is user's behavior, for example 0 indicate clicks, 1 indicate purchase, 2 indicate collections, 3 indicate adding to Shopping cart; visit_time is constituted by month and day. Finally, the form of forecast results is user_id, brand_id1, brand_id2….. and comparing with the actual result of the purchase. The assessment indicators are as follows:

$$\text{precision} = \frac{\sum_i^N hitBrands_i}{\sum_i^N pBrands_i} \tag{1}$$

N is the number of users predicted, pBrandsi is the number of the predicted brand list for the user i, hitBrandsi is the number of intersection between the predicted brand list and really bought brands list for the user i.

$$\text{Recall} = \frac{\sum_i^M hitBrands_i}{\sum_i^M bBrands_i} \tag{2}$$

M is the number of users actually generated transactions, bBrandsi is the number of really bought brands for the user i, hitBrandsi is the number of intersection between the predicted brand list and really bought brands list for the user i.

Finally, F1-Score is used to fit the precision and recall rate.

$$\text{F1} = \frac{2 * \text{P} * \text{R}}{\text{P} + \text{R}} \tag{3}$$

3.1.1 Attribute Selection

This paper constructs 50 property values based on user behavior and date, such as interaction attributes, user attributes, brand attributes and complex attributes. Selection of the property's value play a very important role for the classification recommended of the mass user behavior, good properties can get a better classification results. In the experiment, we divided data set into two parts, the first part is the data set of the first three months as a training set, the other part used as a prediction set, each of which is nearly 90 days of data.

Interaction attribute: interaction attribute that is summarized based on user behavior attributes. As it is a set of training data to predict the final month of the user's purchasing behavior, the closer to the right of the last day of the user's behavior, the more significant. We also make the number of the user clicks, purchase, collection and add to cart behavior with respect to time decay. The coefficient of attenuation is 1 / ((days-1) / 30 + 1), where days is the number of distance from the last day.

Brand attributes: it mainly generate based on the number of this brand's user clicks, purchase, collection and add to the cart.

User attribute: it mainly generate based on the user's own clicks, purchase, collection and add to the cart number.

Composite attribute: it mainly composite interaction attribute, user attributes or brand attributes together (Table 1).

Table 1. The classification of attributes

Attributes	Type	Description
The hits of last 1, 3, 6 days	Interaction attribute	It is mainly based on user clicks and purchases during a period. The selection method is similar with dichotomy period.
The hits of last 7–15, 16–30, 31–60, 61–90 days		
The purchases of last 6, 7–15, 15–30, 31–60, 61–90 days		
The hits of this brand in the last 15, 16–30, 31–60, 60–90 days		It is mainly get hits, the more frequently click this brand, the more interest.
The total numbers of purchase this brand in last 90 days		It is mainly makes statistical sampling based on the number of collection and adding to cart in the last a month. The more number, the more interest.
The number of adding to cart in the last 3, 7, 7–15 days		
The number of collection in the last 7, 15, 30 days		
The days of click this brand in the last 30 days * the days of click this brand in the last 31-60 days		It is mainly used to determine whether the user continued attention or purchase to the brand.
The hits of last 6 days * the hits of last 7–15 days		
(The hits of last 6 days + the hits of last 7–15 days) *(the hits of last 16–30 days)		
(The hits of last 6 days + the hits of last 7–15 days + the hits of last 16–30 days) * sqrt (sqrt(the hits of last 31–60 days))		

(*continued*)

Table 1. (*continued*)

Attributes	Type	Description
(The hits of last 6 days + the hits of last 7–15 days + sqrt(the hits of last 16–30 days)) * (the hits of last 31–60 days) * (the hits of 61–90 days)		
The purchases of last 6 days * the purchases of last 7–15 days		
(The purchases of last 6 days + the purchases of last 7–15 days) * sqrt(the purchases of last 16–30 days))		
(The purchases of last 6 days + the purchases of last 7–15 days + sqrt(the purchases of last 16–30 days)) * sqrt(sqrt(the purchases of last 31–60 days))		
(The purchases of last 6 days + the purchases of last 7–15 days + sqrt(the purchases of last 16–30 days)) * (the purchases of last 31–60 days) * (the purchases of last 61–90 days)		
(The hits of last 6 days + the hits of last 7–15 days) * (the days of click in the last 15 days −1)		
The hits of this brand in the last 1, 3, 7, 15 days/the hits of all		It mainly is a percentage of between attention in the last a pried and total attention. The higher the percentage, the more attention.
The number of knowing and purchase this brand/the total number of knowing this brand (TaoBao conversion rate)	Brand attribute	It is mainly represent the popularity of this brand, smoothly pop or rapidly popular brand should be recommended.
The re-purchase rate of this brand		
The tendency of brand hot (according to the number of purchases)		
Average number of purchase this brand every month		
The purchases of this brand in the last 7 days		
The on-line days of last 10, 20 days	User attribute	It is mainly represent activity status in the near future. The more frequently, the more likely re-purchase.
The days of purchase in the 90 days (frequency)		
The numbers of purchase brand/the numbers of knowing brand		
The purchases of last 3, 7, 15 days/the total purchases		
The days of purchase this brand in the last 30 days * the re-purchases rate of this brand	Composite attribute	It is mainly represented whether users will re-purchase this brand in a month, if users re-purchase this brand, it is likely to purchase this brand more time.

3.1.2 Parameter Configuration

Due to continuous property values, so we can use C4.5 and CART algorithms, in the parameters configuration, the other parameters are the same except the decision tree algorithm. Here, the number of random forest trees is 1000 (range from 10 to 1000), the number of each step algorithm divided attributes is log (N), the maximum number of records per tree is 1,000,000 (range from 1000 to 1000000).

3.2 Training Results

3.2.1 Confusion Matrix

See Table 2.

Table 2. Confusion matrix

	Random forest model based on C4.5	Random forest model based on CART
The Negative examples of correct prediction (TN)	16,444,969	16,449,164
The number of negative examples mistaken positive (FP)	159,915	155,720
The number of positive examples mistaken negative (FN)	1,186,845	1,090,448
The positive examples of correct prediction (TP)	393,039	489,436
The number of actual negative examples	16,604,884	16,604,884
The number of actual positive examples	1,579,884	1,579,884
The number of predicted negative examples	17,631,814	17,539,612
The number of predicted positive examples	552,954	645,156
Total number	18,184,768	18,184,768

3.2.2 ROC Curve

In the Fig. 2, the left is ROC curve of random forest model based on C4.5 decision tree algorithm, the right is ROC curve of random forest model based on CART decision tree algorithm.

The closer to the upper left corner ROC curve, the higher the accuracy of the test. The point closest to the upper left corner of the ROC curve is a minimum fault of best threshold, the total number of false positive and false negative is minimum [13]. As shown in Fig. 2, the ROC curve of random forest model based on CART algorithm is closer to the upper left corner and more accurate.

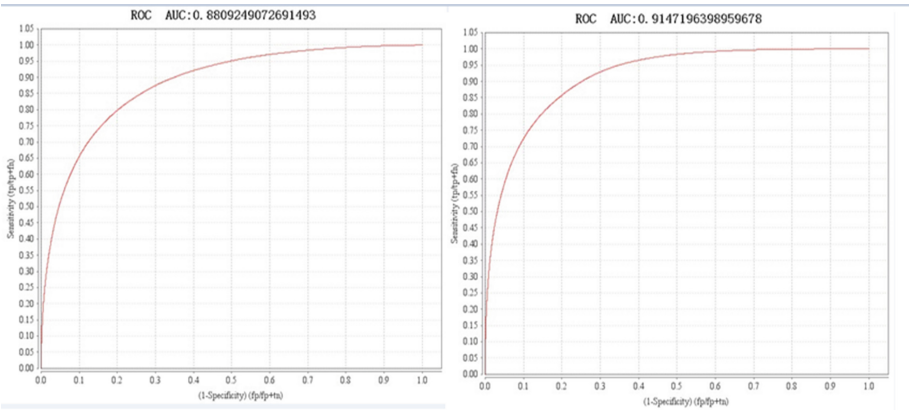


Fig. 2. Curve analysis

3.3 Prediction Results

Based on these two trained models, we predict the brand that the user is about to buy in the next month. Predicted table is mainly constitute by user_id, brand_id, probability where probability is a decimal from 0 to 1 to show the probability users may purchase in next month. As the final validation set is about 2.5 million, we take forecast result that probability are more than 0.4 about 2.8 million data to verification (Table 3).

Table 3. Comparison results of classification

	Precision (P)	Recall rate (R)	F1
Random forest classification result of C4.5	5.66 %	5.64 %	5.65 %
Random forest classification result of CART	5.83 %	5.82 %	5.82 %

The results showed that classification results of random forest based on CART decision tree algorithm is superior to classification results of random forest based on C4.5 decision tree algorithm. Both model evaluation and actual results showed that CART algorithm is better than C4.5 algorithms in user behavior classification decisions.

4 Conclusion

The paper used Random Forest model to classify and compared the results of C4.5 and CART based on the massive actual user data. The results show that CART algorithm is superior to C4.5 algorithm on actual user transactions and difficult attributes will affect the classification results.

References

1. Resnick, P., Varian, H.R.: Recommender systems. Commun. ACM **40**(3), 56–58 (1997)
2. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001)
3. Guo, Y., Liu, X., Guo, M., et al.: Identification of plant resistance gene with random forest. Jisuanji Kexue yu Tansuo **6**(1), 67–77 (2012)
4. Jiangeng, L., Zhikun, G., Xiaogang, R.: Random forest based gene pathway analysis of gastric cancer microarray data. J. Biol. **27**(2), 1–4 (2010)
5. Fang, K., Zhu, J., Xie, B.: A research into the forecasting of fund return rate direction and trading strategies based on the random forest method. Econ. Surv. **2**(9), 61–66 (2010)
6. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques. Elsevier, New York (2011)
7. Srivastava, J., Cooley, R., Deshpande, M., et al.: Web usage mining: discovery and applications of usage patterns from web data. ACM SIGKDD Explor. Newsl. **1**(2), 12–23 (2000)
8. Schapire, R.E.: The strength of weak learnability. Mach. Learn. **5**(2), 197–227 (1990)
9. Freund, Y.: Boosting a weak learning algorithm by majority. Inf. Comput. **121**(2), 256–285 (1995)
10. Quinlan, J.R.: C4.5: Programs for Machine Learning. Elsevier, New York (2014)
11. Breiman, L., Friedman, J., Stone, C.J., et al.: Classification and Regression Trees. CRC Press, Boca Raton (1984)
12. Fan, Y., Chen, L., Qifeng, Z.: Random forest based potential k nearest neighbor classifier and its application in gene expression data. Syst. Eng. Theory Pract. **32**(4), 815–825 (2012)
13. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology **143**(1), 29–36 (1982)

Risk Factors and the Difference Among Hypertension, Diabetes and Heart Disease

Xue Wang[1], Lili He[2], and Hongtao Bai[3(✉)]

[1] Health Promotion Department, China-Japan Union Hospital, Jilin University, Changchun 130012, China
[2] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[3] Center for Computer Fundamental Education, Jilin University, Changchun 130012, China
baihongtao@263.net

Abstract. To investigate the prevalence of hypertension, diabetes and heart disease in the Jilin province of northeast of China and the association with biochemical targets and lifestyles. Subjects and methods: A community-based survey of 5953 adults through 2015 was following a crossing subject on the Chronic Disease Risk Factor Surveillance of China by stepwise. The characteristics are collected by one-to-one interview. Binary logistic regression analysis is conducted to explore the independent factors of the three diseases separately. Results: Prevalence of hypertension, heart disease and diabetes in the northeast of China was 16.1 %, 8.3 % and 5.0 % respectively. Family history and sleep time were independently associated with both hypertension (paternal OR = 2.37, 95 % CI: 1.96–2.79; maternal OR = 1.95, 95 % CI: 1.64–2.31; sleep time OR = 1.24, 95 % CI: 1.11–1.40) and heart disease (paternal OR = 2.20, 95 % CI: 1.72–2.83; maternal OR = 2.82, 95 % CI: 2.25–3.53; sleep time OR = 1.24, 95 % CI: 1.07–1.43). The prevalence of hypertension (OR: 0.68, 95 % CI: 0.55–0.83) and diabetes (OR: 0.67, 95 % CI: 0.50–0.92) in female was lower. Conclusions: Modifiable health risk behaviours are essential for preventing a series of chronic diseases, especially to the people of old age with family history.

Keywords: Hypertension · Diabetes mellitus · Heart disease · Odds ratios · 95 % confidence interval · Binary logistic regression analyses

1 Introduction

Diseases like hypertension, diabetes and heart disease are becoming the dominant cause of death and disability globally. Some patients with single disease always have some risk factors for other chronic diseases. It is important to identify the common and unique independent factors (biochemical targets and lifestyles) to reduce the risk of the three diseases.

During the recent decade, the rate of life style disease that could be characterized as multi-factorial inheritance of risk, such as hypertension, diabetes and heart disease, increased rapidly. The influencing factors of these three chronic diseases were of great differences and similarities. A number of studies have shown that family history of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_73

diabetes (OR = 1.23) was one of the risk factors for nephropathy in diabetic patients [1]. The high prevalence of insulin resistance was observed in the first degree relatives of patients with type 2 diabetes mellitus [2]. There're 54 % of the subjects of hypertension and 28 % of women of heart disease with definite coronary artery disease with a positive family history [3, 4].

It is also important to consider that some modifiable behaviours (physical inactivity, excessive alcohol consumption) and physiological and biochemical targets (blood fat, liver function, electrocardiograph examination, et al.) can increase and indicate the risk of chronic disease. To confirm and adjust some similar and independent factors would be important to both confirmed cases and higher risk populations.

2 Statistical Analysis of Risk Factors

2.1 Statistical Analysis

Database EpiData 3.0 was employed for data input, followed by statistical analysis with SPSS software (version 17.0) to generate descriptive statistics. We investigate the risk factors of the three multifactor diseases based on the analysis of 25 variables. Differences in gender, age, ethnicity, marital status, family history, education, monthly income, BMI, life styles (smoking history, drinking history, exercise, actual sleep time), physiological and biochemical indexes (electrocardiogram (ECG), liver function, blood lipids) were evaluated by chi-square test. A binary logistic regression was conducted to evaluate whether these factors were associated with the incidence of hypertension, diabetes, and heart disease independently by adopting odds ratios (OR) and their corresponding 95 % confidence intervals (95 % CI).

2.2 Risk Factors

A total of 6010 questionnaires were returned from the 26 communities and 5703 ones were retrieved with the effective rate of 96.5 %. Participants, who missed filling out social smoking and drinking status, as well as rejected blood drawing, were excluded from further analysis. The majority of population in this study is Han Chinese (97.2 %). There're 38.2 % participants were female. Mean age was 47.8 $\pm$ 6.8 years. Hypertension ranked top of prevalence of chronic conditions, followed by heart disease (8.5 %) and diabetes (5.6 %).

The prevalence of above diseases was compared by the general characteristics (age, sex, race/ethnicity, nativity, income, education) using chi-square test for the total study population at baseline as showed in Table 1, and then all parameters associated with the three chronic diseases were showed in Table 2 by binary logistic regression analyses.

The significant factors associated with hypertension contained parental inheritance (paternal history OR: 2.37, 95 % CI: 1.96–2.79; maternal history OR:1.95, 95 % CI: 1.64–2.31), lack of exercise (OR: 1.23, 95 % CI: 1.10–1.38), gender (female versus male, OR: 0.68, 95 % CI: 0.55–0.83), sleep less than 5 h (OR: 1.24, 95 % CI: 1.11–1.40), abnormal ECG changes (OR: 1.49, 95 % CI: 1.22–1.81), fatty liver (OR: 1.36,

Table 1. Binary logistic regression analyses are conducted to assess the association between demographic characteristics, life styles, physical indexes and chronic diseases

Variables	Hypertension		Diabetes mellitus		Heart disease	
	OR (95 % CI)	*P*	OR (95 % CI)	*P*	OR (95 % CI)	*P*
Gender	0.68 (0.55–0.83)	<0.01	0.67 (0.50–0.92)	0.01	NS	–
Age	1.98 (1.79–2.19)	<0.01	1.23 (1.06–1.44)	0.01	1.99 [1.75–2.28]	<0.001
Excise	1.23 (1.10–1.38)	<0.01	NS	–	NS	–
BMI	1.72 (1.57–1.89)	<0.01	NS	–	NS	–
Sleeping time	1.24 (1.11–1.40)	<0.01	NS	–	1.24 [1.07–1.43]	0.01
Paternal history	2.37 (1.96–2.79)	<0.01	1.77 (1.21–2.59)	0.01	2.20 [1.72–2.83]	<0.001
Maternal history	1.95 (1.64–2.31)	<0.01	NS	–	2.82 [2.25–3.53]	<0.001
ECG	1.49 (1.22–1.81)	<0.01	NS	0.31	1.81 [1.44–2.27]	<0.001
Fatty liver	1.36 (1.15–1.62)	<0.01	1.39 (1.06–1.81)	0.02	NS	–
GGT	1.43 (1.16–1.76)	<0.01	NS	–	NS	–
TC	NS	–	1.41 (1.01–1.99)	0.05	NS	–
TG	NS	–	2.28 (1.76–2.96)	<0.001	NS	–

95 % CI: 1.15–1.62), older (OR: 1.98, 95 % CI: 1.79–2.19) and high level of GGT (OR: 1.43, 95 % CI: 1.16–1.76).

As for diabetes, fatty liver (OR: 1.39, 95 % CI: 1.06–1.81), high level of CH (OR: 1.41, 95 % CI: 1.01–1.99) and TG (OR: 2.28, 95 % CI: 1.76–2.96), different gender (female with lower risk than male, OR: 0.67, 95 % CI: 0.50–0.92), paternal inheritance (OR: 1.77, 95 % CI: 1.21–2.59) and older (OR: 1.23, 95 % CI: 1.06–1.44) would be liable to contribute to the result.

Besides, taking into account the relative factors, such as parental inheritance (paternal OR: 2.20, 95 % CI: 1.72–2.83 and maternal OR: 2.82, 95 % CI: 2.25–3.53), older (OR:1.99, 95 % CI: 1.75–2.28), sleep less than 5 h (OR: 1.24, 95 % CI: 1.07–1.43) and abnormal changes of ECG (OR: 1.81, 95 % CI: 1.44–2.27), all of them could significantly increase the relevant risk of heart disease.

A most recent report on hypertension prevalence of Chinese came from a full sample of data in 2008 (26.6 %), which had been increased by almost 129.3 % compared with that in1991 (11.6 %) [5]. And in the past 50 years, type 2 diabetes has

emerged as one of the major public health problem based on the reported rate (6.8 %) of the International Diabetes Federation (IDF) Diabetes Atlas in 2013 [6]. Though the result in our study showed a lower prevalence of hypertension (16.1 %) and diabetes (5.6 %) than the recent report, we cannot ignore serious damages caused by the related diseases such as heart disease [7]. It has been difficult to quantify these diseases burden, especially in China with increasing number of senile and pre-senile people. The result may be more serious when multivariate factors coexistent than single risk factor. Hypertension, diabetes and cardiovascular disease are characterized by a combination of several kinds of metabolic disorders, and often coexist with other risk factors, such as family histories, unhealthy lifestyles, and other changed biochemical indexes (Fig. 1).

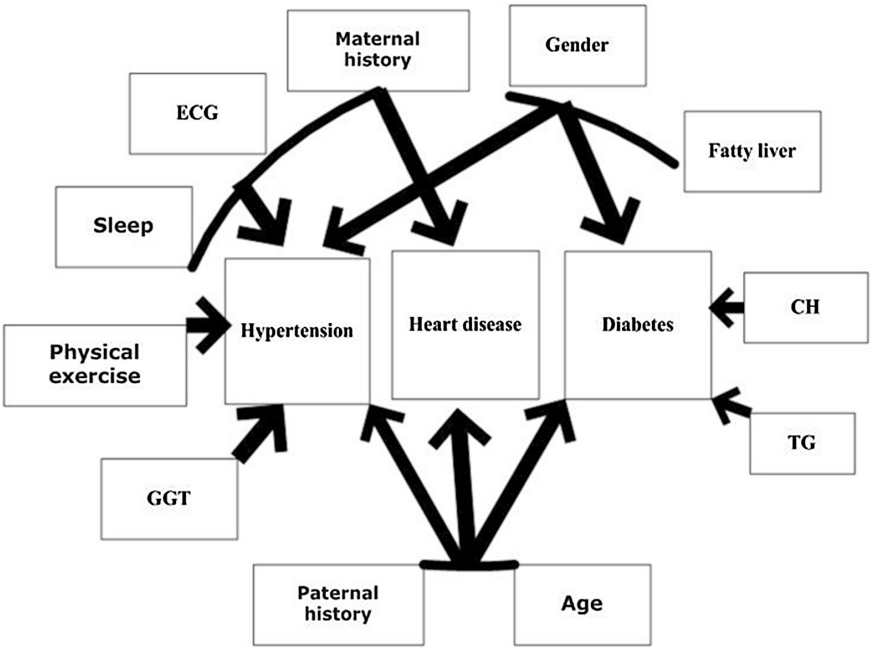


Fig. 1. Independent parameters associated with hypertension, diabetes and heart disease based on binary logistic regression analyses.

OR: odds ratios; 95 % CI: 95 % confidence interval; NS, no significance: variables of characteristics in Table 2 do not exactly match those in Table 1, which were derived from a competing risks analysis which allowed the variables to change based on who was still "at risk" of screening.P values calculated as $p = 0.000$ by the software spss 17.0 were reported as "$P < 0.001$".

3 Conclusion

The incidences of the three diseases were related to the genetic history and increased with age. The data suggest that the male and the person with fatty liver will face a significantly higher risk of developing hypertension and diabetes. Sleep less than 5 h every day would make a person under higher risk for both hypertension and heart attacks, and the person with abnormal changes of ECG will develop not only heart disease but also hypertension. The serum GGT level was independently and specifically associated with hypertensive patients rather than diabetes and cardiac patient.

Some patients with single disease always have some risk factors for other chronic diseases [8, 9]. It is important to identify the common and unique independent factors for the people with some diagnosed or undiagnosed chronic diseases [10–12].

Though the data has no classification (in this cross-sectional study, without the grade of hypertension, the classification of diabetes and no checks on heart disease), given the independent and common risk factors of hypertension, diabetes and heart disease, understanding the actions (regular exercise, exercise enhancement, moderation sleep and health examination management) that people at high risk can take to ameliorate this risk is essential in an ageing society.

Acknowledgements. The authors gratefully acknowledge the financial supports for this research from the National Natural Science Foundation of China (51409117, 61272208, 61572228), Jilin Province Department of Education "Thirteen Five" scientific and technological research projects [2016] No. 432, Fundamental Research Funds for the Central Universities Special (JCKY-QKJC47, JCKY-QKJC49).

References

1. Ahmed, M.A., Kishore, G., Khader, H.A., et al.: Risk factors and management of diabetic nephropathy. Saudi J. Kidney Dis. Transplant. **24**(6), 1242 (2013)
2. Jyothirmayi, B., Kumar, J.S.: Insulin resistance and alanine amino transaminase (ALT) levels in first degree relatives of type 2 diabetes mellitus. Diabetes Metab. Syndrome: Clin. Res. Rev. **5**(3), 143–147 (2011)
3. Babiker, F.A., Elkhalifa Lamia, A., Moukhyer, M.E.: Awareness of hypertension and factors associated with uncontrolled hypertension in sudanese adults. Cardiovasc. J. Africa **24**(6), 208 (2013)
4. Salehi, R., Motemavele, M., Goldust, M.: Risk factors of coronary artery disease in women. Pak. J. Biol. Sci. **16**(4), 195 (2013)
5. Wu, X., Duan, X., Gu, D., et al.: Prevalence of hypertension and its trends in Chinese populations. Int. J. Cardiol. **52**(1), 39–44 (1995)
6. Gao, Y., Chen, G., Tian, H., et al.: Prevalence of hypertension in China: a cross-sectional study. PLoS ONE **8**(6), e65938 (2013)
7. Chan, J.C.N., Cho, N.H., Tajima, N., et al.: Diabetes in the western pacific region—past, present and future. Diabetes Res. Clin. Pract. **103**(2), 244–255 (2014)
8. Park, J.B., Kario, K., Wang, J.G.: Systolic hypertension: an increasing clinical challenge in Asia. Hyperten. Res. **38**(4), 227–236 (2015)

9. Krittayaphong, R., Rangsin, R., Thinkhamrop, B., et al.: Prevalence and associating factors of atrial fibrillation in patients with hypertension: a nation-wide study. BMC Cardiovasc. Disord. **16**(1), 1 (2016)
10. Kim, I.G., So, W.Y., Sung, D.J.: The relationships between lifestyle factors and hypertension in community-dwelling Korean adults. J. Phys. Therapy Sci. **27**(12), 3689–3692 (2015)
11. Angeli, F., Angeli, E., Verdecchia, P.: Novel electrocardiographic patterns for the prediction of hypertensive disorders of pregnancy—from pathophysiology to practical implications. Int. J. Mol. Sci. **16**(8), 18454–18473 (2015)
12. Eleftheriou, P., Tseka, E., Varaga, E., et al.: Study of the lipidemic profile of diabetic patients. Negative correlation of cholesterol levels of diabetes type I patients with serum amylase concentration. Hellenic J. Nucl. Med. **17**, 35–39 (2013)

Linear Programming Computation Model Based on DPVM

Hongtao Bai[1(✉)], Lili He[2], Yu Jiang[2], Jin Wang[3], and Shanshan Jiang[2]

[1] Center for Computer Fundamental Education, Jilin University, Changchun 130012, China
baihongtao@263.net
[2] College of Computer Science and Technology, Jilin University, Changchun 130012, China
[3] Information Engineering College, Yangzhou University, Yangzhou 225009, China

Abstract. Matrix manipulation of Linear Programming (LP) problems is a performance bottleneck in Single Instruction Single Data (SIMD) pattern. While, GPU is specialized for this compute-intensive and highly parallel computation, which is exactly what graphics rendering is about, due to the Single Instruction Multiple Data (SIMD) architecture. This paper introduces a Revised Simplex Method (RSM) on a GPU–Data Parallel Virtual Machine (DPVM). It assigns different routines for CPU and GPU according to respective characteristics: Iteration control and minimum value obtained are completed by CPU and Matrix multiplication by DPVM. In detail, we carefully organize the data as 4-channel textures, and efficiently implement the computation using Fetch4 technology of pixel shader. Numerical experiments are presented to verify the practical value and performance of this algorithm. The results are very promising. In particular, they reveal that our method not only can get correct optimal solution, but also is sixty-six faster than a traditional method on CPU, near 2.5 times faster than a lasted released MATLAB when LP problem size reaches 1200.

Keywords: Revised simplex method · GPU · DPVM · Pixel shader

1 Introduction

Revised Simplex Method (RSM) is the most representative and widely algorithms used for solving Linear Programming (LP) problems. Solving LP methods are mainly divided into two categories: simplex category [1] and interior point type [2]. Though the worst time complexity of Simplex Method is exponential level [3], the method has been extensively studied owing to good performance in the small and medium-sized real problem. Recently, a new method combined outer point simplex, interior point and simplex method to avoid the feasible solution has proposed by Paparrizos et al. [4]; Luh and Tsaih put an alternative method replaced the simplex methodwith interior point [5]; Junior and Lins raised cosine iterations method to get better initial feasible

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_74

solution [6], and so on. These methods are to improve efficiencies by reducing the number of iterations of the simplex method.

With the increasing scale of the practical problems, the efficiencies become an increasingly important indicator. Foreign molding solving systems for optimization problem such as MATLAB, CPLEX and LINDO/LINGO, etc. begin with customary efficiency, and with the expansion of the scale of the problem, its running time inevitably increase substantially. Matrix representation of LP problems and a series of solving operations with computationally intensive, parallel features in the CPU, so serial execution is obviously not the best choice. Using SIMD (SIMD) multiprocessor systems parallel computing, is expected to achieve high computational efficiency than the CPU.

Today, the large numbers of SIMD processors constitute GPU (Graphics Processing Unit, referred to as the GPU) as standard desktop computers. Traditionally, the main task of GPU is to graphics drawing and image rendering. With opening of the programmable interface and popularity of advanced rendering language, GPU with its powerful parallel processing capability and high memory bandwidth extensions to the general computation. Good results have achieved in several aspects, such as: type identification [7], wavelet transform [8], computational intelligence [9], and real-time simulation [10], and so on. GPU become a most cost-effective branch in high-performance computing.

Based on the general-purpose processor platform DPVM (Data Parallel Virtual Machine) [11], proposed revised simplex method which is suitable for the GPU (GPU-based RSM) is proposed in this paper. Making full use the advantages of flexible CPU and computing-intensive advantages GPU, CPU's task is to take control of the minimum and iterative work, and GPU's task is to perform calculations related to matrix operations. Previously open a GPU texture space needed, design efficient data storage and transmission mode, and organize tissue texture by 4-channel, and use Fetch4 parallel computing technology. Numerical test analysis shows that the algorithm is correct and effective. When the scale of the problem reaches 1200, the method for solving the same problem is 66 times faster than the CPU version (CPU-based RSM), and faster than MATLAB nearly 2.5 times.

2 Modified Simplex Method Based on DPVM

GPU's SIMD architecture means that at any one clock cycle, each single-processor executes the same instruction, but operates on different data. From this, you can put among a number of different data manipulations on large data sets independent calculations to the GPU to complete. GPU computing has the advantage of such competition in following two aspects:

(1) Floating point computing power of the GPU and CPU disparities largely. Firestream 2U used in this paper can handle 375 GFLOPS (giga floating-point operations per second), while an Intel Core2 Duo 3.0 GHz of CPU can only reach 48 GFLOPS.

(2) GPU has a higher memory bandwidth than the system's main CPU. Intel felt pride once for Pentium 4 XE series processors have FSB 1066 MHz provided 8.6 GB/S bandwidth, while Firestream 2U owns 160 GB/S or more memory bandwidth.

Revised simplex method is the basic method for solving LP problems, and is widely used because of its good and suitable for solving the structure on the computer. The most basic method is the most time-consuming calculations - Matrix operations, which become a performance bottleneck. Noting $C = A \times B$ (Fig. 1), any element of computing a matrix is independence, $c_{i,j} = \sum_{l=1}^{m} a_{i,l} * b_{l,j}$. That is, their calculation rules are exactly the same, act on different data, and have parallel features, which is that the GPU SIMD architecture is good at.

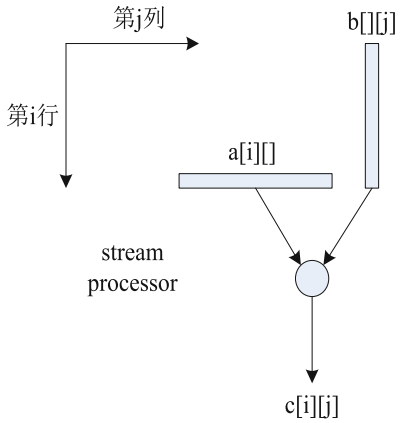


Fig. 1. Example of matrix multiplication

Since the flow model architecture of GPU, there are still many restrictions in GPU general computing, and the traditional algorithm on the CPU simply ported to the GPU is not a good play to its performance. First, GPU cannot self-start and I/O, and must rely on the CPU environment settings and data, code loading, and other initialization task only from the processor as a CPU work. Secondly, GPU vast majority of transistors are calculating unit (ALU), so its powerful computing power and the flow control is weak. Finally, the GPU general computing programmable interface is not perfect and there are still many constraints, or to do special processing for the general-purpose computing. Thus, we use the powerful data processing capabilities of GPU, on the other hand with the advantages of CPU to make up for the d disadvantages of the GPU.

GPU's work is done to solve permanent of compute-intensive associated with matrix, such as: the simplex multipliers, test number δ, ratio θ, etc. These calculations can take full advantage of GPU parallel flow model; control of iteration, swapping into variable and swapping out variable and other operations have been assigned to the CPU execution. Exchange variables are to select the minimum element in a sequence. If the

work of this part is implemented on GPU, there are two alternative methods: one is the pixel program starts only one pixel (pixel) to calculate the task, which will inevitably result in the rest of the processor is idle, and GPU processor execution will be evaluated within a large overhead in a loop, far less than the efficiency of CPU; The second is to selected the minimum value based on the GPU parallel sorting, and the current time complexity of the most efficient sorting algorithm for the GPU is $O(n \overset{2}{\log} n)$. For DPVM experimental tests also show that these two methods take longer time than the CPU.

CPU memory, shared memory and data transfer between the memories bring extra overhead GPU in general computing. In this paper, the data transfer occurs in the calculation model: (1) At the data initialization, the coefficient matrix A, resource vector and vector values b, c uploaded to the memory at once; (2) Ratios sequence of test number δ to θ download to memory; (3) The parameters x_B, c_B for calculating the optimal solution return to memory once after the end of the iteration. Wherein (1) and (3) occurs only once in the entire solution, and it does not affect efficiency of the algorithm. Though (2) occurs during iteration, (2) are 1D vector with $O(n)$ spatial complexity. Overall, the model minimizes the amount of data transfer.

Pixel program memory access speed is much higher than the speed of its access to shared memory, only to maximize the use of memory in order to enhance the efficiency of the algorithm and meet the requirements of large-scale problem solving. GPU used in this paper is in addition to 831 MB of memory, but also has 96 MB of shared memory resources. Thus, the design data storage mode as: The matrix A, b, c is read into memory by the CPU, and the CPU is transferred to the shared memory, and then upload by the pixel program and reside in memory static texture space; and the ratio sequences of the test number δ and θ stores in the shared memory; intermediate data such as: E, A_p, T_p and so on in the iterative algorithm, apply and release static texture space directly on the memory; results, download the program from memory by the pixel static texture space to shared memory, the CPU and then copied back to memory from the shared memory. Results x_B, c_B downloaded by the pixel program from texture space to a static memory shared memory, and then CPU copies back to memory from the shared memory. This will not only make full use of the memory capacity and bandwidth advantages, but also to avoid frequent data transmission of the memory to the shared memory. In addition, the shared memory and memory in the float4 form usage of four texture channels <r, g, b, a>, the data as the linear data and packaging data ways fold into a two-dimensional matrix texture. Linear folding places texels as mapping units, and four channels each texel occupy the four value of the original data stored in connected order to support matrix division and subtraction operations; data package is folded to 2×2 sub-matrix form and add to texture, and texel elements are in the four channels by block matrices order to support the matrix multiplication operation, the detailed structure and operation can refer to the literature [12]. Pixel process has the same clock compered completed in one 4D vector operation to one 1D operation. The article with the data storage mode, use Fetch4 function, first read the four values from the texture, all the basic operations are in 4D vector units. Storage space utilization and efficiency while increasing four times. In this paper, use Fetch4 function and read the four values from the texture once with the data storage mode, and

all the basic operations are in 4D vector units. Utilization of storage space and efficiency of completing increases four times meanwhile.

3 Experiment Analysis

First, we test the GPU-based RSM can get or not the correct optimization results. Experimental platform is the Intel Pentium 2.8 GHz, the system main memory is 2 GB, GPU is ATI FireStream 2U, and available memory is 831 MB. Test samples generated by the CPU, and in order to get a more accurate comparison, the data taken randomly generate 32-bit single-precision floating-point in interval [0,1]. Setting the standard LP problem:

$$\max z = c^T x \quad s.t. \begin{cases} Ax = b \\ x \geq 0 \end{cases}$$

$A \in R^{m \times n}$, $b \in R^m$, $c \in R^n$. The number of constraints is same as the number of original variables, that is, $n = 2m$, and the number of slack variables is m. Linear optimization software MATLAB, CPU-based RSM and GPU-based RSM are used for solving, as shown in Table 1:

Table 1. Correctness test

LPs size (m)	MATLAB	CPU-based RSM		GPU-based RSM	
	Optimal value	Iterations	Optimal value	Iterations	Optimal value
100	0.039	8	0.039	8	0.039
200	0.065	13	0.065	13	0.065
300	0.012	13	0.012	13	0.012
500	0.026	7	0.026	7	0.026
600	0.020	8	0.020	8	0.020
800	0.009	11	0.009	11	0.009
1000	0.007	10	0.007	10	0.007
1200	0.010	24	0.010	24	0.010

As can be seen from the table, three methods defined under the effective number of bits (three decimal places) precision get the same optimization results in all test samples. In addition, GPU-based RSM also has the same number of iterations with CPU-based RSM. MATLAB algorithm will be depending on the choice of the size and characteristics of the different LP the number of iterations is not as baseline. Thus, though the CPU and GPU different floating-point precision, DPVM 32-bit floating-point number has not yet reached the IEEE-754 standard, GPU still can be applied to optimize the calculation. So the method proposed is effective.

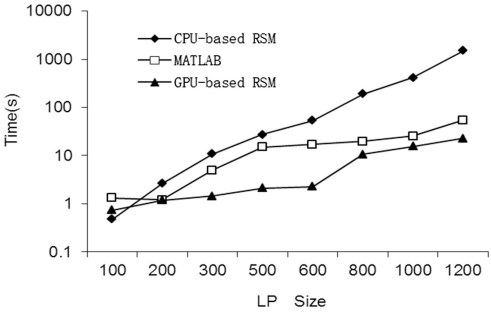


Fig. 2. Efficiency compared

Secondly, in order to verify the efficiency of the algorithm, Generate 20 the same scale compute instance of the problem, and take the average running time. The results are shown in Fig. 2:

When the scale of the problem is relatively small and amount of computation is intensive, data transfer between main memory and the memory during the entire solving is too large ratio, making efficiency of the GPU-based RSM less than MATLAB and CPU-based RSM. When the scale becomes larger, the performance benefits of GPU gradually appear. Although MATLAB make a lot of low-level optimization for CPU, such as real-time JIT compiler and techniques of vectorization and so on, resulting in much faster than CPU code that is written in accordance with the C/C ++ standard, GPU-based RSM is faster nearly 2.5 times than MATLAB and faster 66 times than CPU-based RSM when the scale of the problem reaches 1200.

4 Conclusion

A revised simplex method suited to GPU is proposed in this paper based on GPU platform DPVM architecture. Give a full play to their strengths by dividing different roles: the task of CPU is only for control and comparison operations, and task of GPU is for complex floating-point operations. In the solution process, reduce the data transfer between the CPU and GPU by large-scale data residing in the GPU memory, and accelerate the entire calculation process. Numerical experiments showed that involvement of GPU improve the efficiency of the algorithm greatly in premise of ensuring the accuracy. Further studies will focus on how to make breakthrough in a certain limitations of GPU resources for solving linear programming problems on a larger scale.

Acknowledgements. The authors gratefully acknowledge the financial supports for this research from the National Natural Science Foundation of China (51409117, 61272208, 61572228), Jilin Province Department of Education “Thirteen Five” scientific and technological research projects [2016] No. 432, Fundamental Research Funds for the Central Universities Special (JCKY-QKJC47, JCKY-QKJC49).

References

1. Pan, P.Q.: A primal deficient-basis simplex algorithm for linear programming. Appl. Math. Comput. **196**(2), 898–912 (2008)
2. Kebbiche, Z., Keraghel, A., Yassine, A.: Extension of a projective interior point method for linearly constrained convex programming. Appl. Math. Comput. **193**(2), 553–559 (2007)
3. Klee, V., Minty, G.J.: How good is the simplex algorithm. Washington University of Seattle, Department of Mathematics (1970)
4. Paparrizos, K., Samaras, N., Stephanides, G.: A new efficient primal dual simplex algorithm. Comput. Oper. Res. **30**(9), 1383–1399 (2003)
5. Luh, H., Tsaih, R.: An efficient search direction for linear programming problems. Comput. Oper. Res. **29**(2), 195–203 (2002)
6. Junior, H.V., Lins, M.P.E.: An improved initial basis for the simplex algorithm. Comput. Oper. Res. **32**(8), 1983–1993 (2005)
7. Napoli, C., Pappalardo, G., Tramontana, E., et al.: A cloud-distributed gpu architecture for pattern identification in segmented detectors big-data surveys. Comput. J. 59(3), 338–352 (2014). doi:10.1093/comjnl/bxu147
8. Georgieva, K., Koch, C., König, M.: Wavelet transform on Multi-GPU for real-time pavement distress detection. In: Computing in Civil Engineering 2015, pp. 99–106. ASCE (2015)
9. Ting, T.O., Ma, J., Kim, K.S., et al.: Multicores and GPU utilization in parallel swarm algorithm for parameter estimation of photovoltaic cell model. Appl. Soft Comput. **40**, 58–63 (2016)
10. Conti, F., Khatib, O.: A framework for real-time multi-contact multi-body dynamic simulation. In: Inaba, M., Corke, P. (eds.) Robotics Research. STAR, vol. 114, pp. 271–287. Springer, Heidelberg (2016). doi:10.1007/978-3-319-28872-7_16
11. Peercy, M., Segal, M., Gerstmann, D.: A performance-oriented data parallel virtual machine for GPUs. In: ACM SIGGRAPH 2006 Sketches, p. 184. ACM (2006)
12. Lefohn, A., Kniss, J., Owens, J.: Implementing efficient parallel data structures on GPUs. GPU Gems **2**, 521–545 (2005)

A Study on the Effective Communication Protocol of the Surface Inspection Rail Robot that it can be a Self-checking

Yun-Seok Lee[1], Eun Kim[1], Sungyun Kim[1], and Seokhoon Kim[2(✉)]

[1] LK Smart Co., Seoul, Republic of Korea
{lysis2jt, sil7777, sungyunkim}@lksmart.co.kr
[2] Department of Computer Software Engineering,
Soonchunhyang University, Asan, Republic of Korea
seokhoon@sch.ac.kr

Abstract. The robots that they inspect the rail surface with running above the rail will be a means to replace the existing expensive equipments. This can be very effectively used in the process of construction and maintenance of continuous increased rail industry. Especially, it can detect the small change via the build of database for detecting the state of the rail. And it is possible the continuous management by recording this information. However, these rail robots are attached a number of different rail parts. Thus, if these components are not gonna work, it can occur problems such as stop on the rail or derailed. These issues can be occurred to other train that it runs above the rail. Therefore, it needs the method for ensuring the safety. In this paper, in order to solve these problems, we define the components and performance parts that rail robot should have it. And also, we divide into three groups of devices by analyzing the properties of each of the components. Each of the group was divided according to the calculation processing capability and communication capability. The self-verification of equipments performs with fitting the framework. And they examine its own equipment validation and review whether the operation is enabled or impossible to perform in accordance with the information. In addition, the rail robot continues send the information for the rail status via wireless network with control center. Thus, we propose an efficient communication protocol in consideration of security issues that there may occur in sending process and the scalability of the function side for rail robot.

Keywords: Rail robot · Communication protocol · Rail inpection · Self-checking

1 Introduction

The importance of the environment is gradually increasing. Thus, there is also increasing interest in the efficiency of the vehicle [1–7]. It becomes an opportunity to be moved by rail from the transport that has been many advanced around the car. Nowadays, the world rail market is progressed the large-scale railway projects all over the world including the China, India. According to the German firm SCI verkehr as the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_75

railway statistics professional organizations, world rail market is expected to reach € 190 billion by growing and annual average of 3.4 % by 2018. The broad and long railway is need for the continued growth like this [1]. The railway is need sustainable management for that vehicle runs efficiently and safely. To this, there is a way to determine directly the eye from person, how to use manual inspection methods by directly pulling the inspection equipment, and a way to inspect by more than three people on board inspection vehicle in earlier method [1, 2, 4]. Firstly, how to check by people with their eyes have disadvantage that may be reduced the accuracy. And to inspect by using a manual inspection device has also disadvantage that it is very slow. Finally, by using the inspection vehicle with running on the route is very expensive. And this method disadvantage that a lot of people need to inspect. In order to solve these problems, the study about rail robot that it can effectively inspect with running above the railway is currently proceeding. The rail robot with running above the railway can effectively verify the surface cracks, a distortional rail and a ground subsidence [1]. And they can safely store by the database for the corresponding content. And also they can report information about the rail status after comparing and verifying by loading previously stored data. The rail robot with having these features is relatively inexpensive and can be operated in an unattended station. It can reduce the cost of human and time because it can fast inspect relatively compared with a manual inspection method [1]. However, the biggest disadvantage of this rail robot is that it is difficult to cope when a problem arises, because of no involvement of the people [1, 2, 4]. The rail robot has the various types of components and corresponding devices make determination by the overall role of oneself. And finally, they also make a process the control for operation. At this time, the rail robot can be inoperative by one component causes the problem. In this case, there is also affect to other vehicles with running above the rail. This is not simply caused only damage of the rail robot. It can cause an accident that the freight trains or the train with carrying a number of passengers may be in conflict with the rail robot. In order to solve these problems, it is required by precision self-checking primarily for the rail robots. In addition, to make a safe exchange of information with control center, it need a self-checking protocol of the rail robot and the information exchange protocol. In this paper, we divide and define according to the device performance and finally, also define the required contents of the device for this purpose. In addition, we propose a protocol for effective inspections and communication with the control center.

2 Our Proposed Device

In this paper, we classify the available device, inspection methods and self-checking communication protocol for it that robot has device above the rail way. Also we propose a method that it is possible an effective control through the effective protocol with management center.

A. The device type of the rail robot

The rail robot operates by driving the wheels. Therefore, it requires a drive controller for driving the motor. And the robot must be able to determine their current location.

Thus, it is also required GPS for checking its own position. It is also necessary a surveillance camera. Also to check the status of the robot or the rail, it is also needed to obtain and manage the information for the sensors because various sensors are necessary. And it is also necessary that effectively send management center the information of the robot via the network. And finally, it is required the controller to control the all devices. This controller should have a processing capacity of a high performance. Such hardware configuration is the most basic configuration. Thus, we can say that it is an essential element for operating the corresponding robot. Looking at such devices, each device should examine whether it has the capable of processing for the data. And it must be checked whether it can send processed data to control unit through the communication. In this paper, we classify devices with a total of three types on the basis of these criteria (Table 1).

Table 1. The classification by device type

Type	Title	Description
A	Rail robot central control computer	A computer that controls the whole devices
A	Management system	A computer that monitors the rail robot
B	Rail surface image processing device	A device that inspects the rail surface
B	Drive controller	It is connected to the motor part with a small processor to proceed to remove the drive
B	Communication controller	A part with a small process that controls data when the communication like a 3G/LTE or WIFI
C	Motor	A motor for control the wheels
C	GPS	A module for catching the position
C	Sensor	A sensor for checking the rail status and the attitude control
C	Mars light	A device for informing the status that robot is operating
C	Forward looking device	A device that checks the front of robot and sends the image to controller

B. Self-Checking Method by the type of devices

As previously confirmed, devices are divided according to the processing capacity. Thus, each of the devices are different the checking ways by type. Firstly, inspection method of the type C is that by a periodic sending a message RDR (Response message of check Device Result). The message RDR is a response data that is periodically sent as like a GPS, a Sensor, a front inspection device and a motor. At this time, the mars lights only performs ON/OFF, and it does not send the response data. Thus, its operating should be visually checked with the naked eye (Fig. 1).

The type B is conducted in accordance with the request message for inspection and its response. At this time, if the RDC (Request message of Device Check) such as request message for inspection is sent, it sends the RDR as response for its request (Fig. 2).

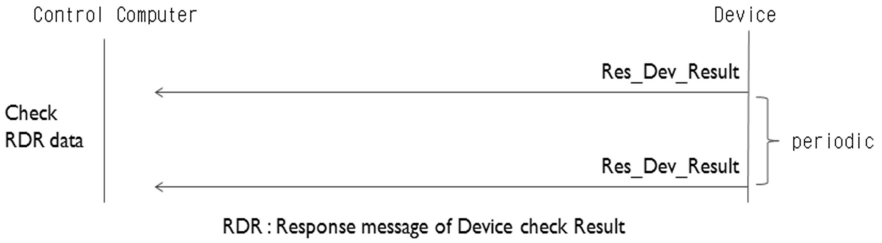


Fig. 1. Inspection method of type C for device

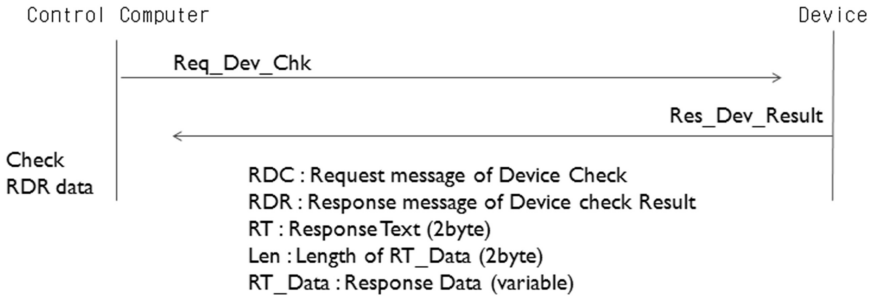


Fig. 2. Inspection method of type B for device

The RDR is as follows.

$$RDR = \{RT \| Len \| RT\_Data\}$$

The normal criteria are shown in the following table (Table 2).

Table 2. Driving Part RT_Data

RT_Data	Status
0000	Normal state
0001	Failure with front right wheel
0010	Failure with front left wheel
0100	Failure with rear right wheel
1000	Failure with real left wheel

If all the individual RT_Data is 0000, it is a normal state. However, all value of the other is all abnormal conditions. At this point, we organize the code as above values for the verification of the state. Thus, even if there is a problem in the front left and right wheels of the driving unit, code value is became the 0001 + 0010 = 0011. Therefore, it can be checked it has a problem in a front left and right wheels. Finally, if it does not be progressed the inspection procedure for the type B, C, we should be expected to be a problem for the inspection of the type A. In case of network state, inspection procedure for the type A is very simple, because it can be verified through the Packet loss by Ping (Table 3).

Table 3. Rail surface image processor RT_Data

RT_Data	Status
0000	Normal state
0001	Failure with right image processor
0010	Failure with left image processor
0100	Abnormal temperature
1000	Camera disruption

C. Control Protocol

The control protocol is the protocol that it should be performed the real control, because it is not an inspection. Therefore, it can be divided into two categories such as the control computer sends the control command to each of the devices and the monitoring system sends the control command to the control computer. In each case, it is a default that it is sent response for the control command (Fig. 3).

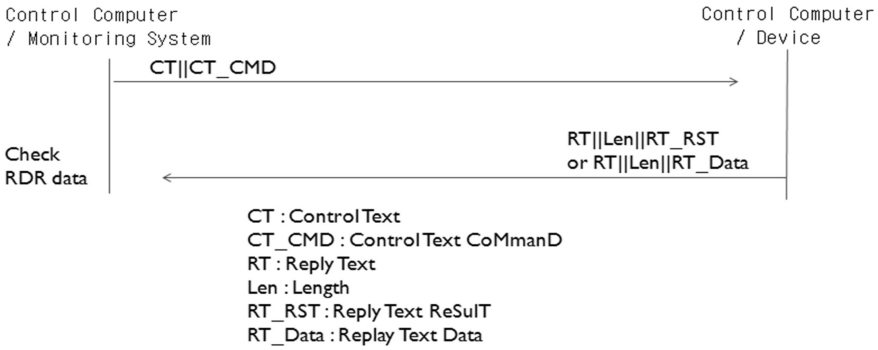


Fig. 3. Control protocol

The Control protocol is given on the basis of sending of the CT and CT_CMD and for receiving the response for this. The response is as follows.

$$Response\,Message = \{RT||Len||(RT\_RST/RT\_Data)\}$$

At this time, the RT_RST is described as indicating information for that a particular status or condition with defined state information in advance. This part is very different for each device. And RT_Data can send and receive the long length data as response data. Like this, the control protocol has a very simple structure of TLV (Tag, Length, Value) type.

3 Conclusion

Today, the rail robot is designed to operate a specific interval. Thus, it has been developed for various purpose because of an advantage that it clearly facilitates the development and management. However, the biggest problem of studies for these rail robot is that it can not be performed a validation of the rail state. And also, there are not present a common inspection procedure and the protocol for rail robot that it is composed with several components. In this paper, in order to solve these problems, we composed a part which is responsible for the image processing module. And we designed an inspection procedure for this. And also, we propose to design the protocol for effective control. The rail robot that it will be developed in the future can directly verify the status of the rails through this. And we proposed the inspection procedure and classification method for at least components of rail robot. Thus, we provide the efficiency for the development process.

Acknowledgments. This paper has received the support of the Research and Business Development (R&BD), Ministry of Trade, Republic of Korea (C0332683).

References

1. Lee, Y.S., Kim, E., Kim, Y.B., Kim, J.M., Kim, Y.S.: Rail inspection robot using the AHRS algorithms. IEER (2016)
2. Li, Q., Zhong, Z., Liang, Z., Liang, Y.: Rail inspection meets big data: methods and trends. In: International Conference on Network-Based Information Systems, pp. 302–308 (2015)
3. Papaelias, M.P., Roberts, C., Davis, C.L.: A review on non-destructive evaluation of rails: state-of-the-art and future development. Proc. Inst. Mech. Eng. F, J. Rail Rapid Transit **222**, 367–384 (2008)
4. Li, Y., Trinh, H., Haas, N., Otto, C., Pankanti, S.: Rail component detection, optimization, and assessment for automatic rail track inspection. IEEE Trans. Intell. Transp. Syst. **15**, 760–770 (2013)
5. Chen, Y., Qian, H., Xu, Y.: Optimization for rail-type climbing robot in space. In: IEEE International Conference on Mechatronics and Automation, pp. 1540–1545 (2013)
6. Moon, S.M., Hong, D., Kim, S.W., Park, S.: Building wall maintenance robot based on built-in guide rail. In: ICIT, pp. 498–503 (2012)
7. Liu, Z., Li, W., Xue, F., Xiafang, J., Bu, B., Yi, Z.: Eletromagnetic tomography rail defect inspection. IEEE Trans. Magnet. **51** (2015)

The Efficient Multimedia Transmission Services for the E-learning System with Sensor

Sung-Hwa Hong[1] and Joon-Min Gil[2(✉)]

[1] Department of Maritime Information and Communication Engineering, Mokpo National Maritime University, 91 Haeyangdaehak-ro, Chukkyo-dong, Mokpo, Jeonnam 58628, South Korea
shhong@mmu.ac.kr

[2] School of Information Technology Engineering, Catholic University of Daegu, 13-13 Hayang-ro, Hayang-eup, Gyeongsan, Gyeongbuk 38430, South Korea
jmgil@cu.ac.kr

Abstract. In this paper, we define and implement the protocol to manage networks in USN (ubiquitous sensor network). Although network management system in USN related with this paper is being progressed for the purpose of independent projects, protocol interfaces and message systems have not clearly defined yet such as general-purpose network or extension into heterogeneous kinds, communication support, etc. Therefore, USN network management should be conducted for the management of fault, composition, power, and applications. In this paper, we design the ubiquitous sensor network based on u-learning processing for multiple devices in order to connect other devices for the multimedia services, allowing the characteristics of existing network protocols.

Keywords: Multimedia · Sensor networks · E-learning system · USN · Protocol

1 Introduction

Recently, researches on ubiquitous computing environment are going on in a variety of fields with appearance of ubiquitous paradigm, and to realize those environments some required technologies should be prepared. In this respect, sensor network technologies will exert an important role. Users can get needed services anytime and anywhere through the actions of physical space and subjects in an intelligent way with a construction of network of their own. Presently, those technologies are being applied to various fields such as industries, environments, and medicine, and will be expanded to a region within our daily lives. Recently, the popularity of portable computers combined with the growth of wireless networks and services has led to many efforts to make mobile computing every reality.

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2055463).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_76

Various kinds of applications and protocols are running and developing to run on the Internet. A new protocol or application after being developed has to be tested on the actual Internet or simulated on a testbed for debugging and performance evaluation. Since Internet is a large and complex network, it is not practical to conduct sufficient experiments directly on the real network. Up to now, the most auspicious way to realize the seamless handover is to store the packets at AP as a provision for the dropping of packets during the handover.

Wireless networks usually consist of a wired, packet-switched, backbone network and one or more wireless hops connecting WMN (Wireless Mobile Node) to wired part. The wireless part is organized into geographically defined cell controlled by AP (Access Point) for each of these cells. These APs are on the wired network and provide a gateway for communication between the wireless infrastructure and the backbone interconnect. These wireless networks give the computer the ability of moving from one domain to another easily. However, this causes the problem how to serve a seamless network connection to a mobile node regardless of its attachment to some domain [1–4].

The current ubiquitous sensor network is wireless network where the sensor nodes of compact size with low cost and power are randomly arranged and composed in a wide area to gather environment-related information. Though USN operates in form of existing wireless ad-hoc networks, it has technically different characteristics from wireless ad-hoc network in terms of network scale, energy efficiency, limited node resources, etc. While existing wireless ad-hoc networks have been built for the purpose of providing communication services between mobile nodes even in a condition without communication infrastructures, USNs are built to gather environment-related information. Thus, its important factors are rather the subject of sensing and the mobility of sink than the mobility of nodes, because such factors require frequent flooding or rerouting to gather sensing information, resulting in consumption of numerous energy sources of USN.

It is very important to reflect these characteristics in designing the network protocol of USN. However, most of existing studies propose plans to gather a range of home sensing information but do not allow for the subject of sensing and the mobility of sink [4–6]. Therefore, the network layer of USN requires network protocol considering the subject of sensing and the mobility of sink while minimizing energy consumption. USN having this feature is specifically suited to obtain and utilize information around room. Information around room can measure various environmental changes around room by utilizing wireless sensor network composed of small-size sensor nodes where sensing, data processing, and communication are enabled. Data measured from the sensor nodes are transferred to the server having the function of gathering, processing, and delivering all data to the user. This process should consider sensor nodes with energy limit.

Thanks to the information technology, the e-learning technology is evolving gradually to a form of the u-learning while the ubiquitous environment is accelerating. It is expected that the e-learning supportive platform which can be used for studying using the terminal (cellular phone, PDA, laptop computer, PMP, etc.) at anytime, anywhere will be appeared in the market.

In this paper, the ubiquitous sensor network based on u-learning processing for multiple devices was designed in order to connect other devices for the multimedia services, allowing for such characteristics of existing network protocol.

2 USN and U-learning System

2.1 USN

Sensor network suffers frequent changes in network topology on account of battery shortage of sensor nodes, disconnected wireless link, addition of a new sensor node, movement of sensor nodes, etc. These changes should be monitored in real time, and sensor network should be designed to manage network topology by independently coping with network environment with automatic network restoration function. Also, such changes should be monitored all the time and be displayed in an administrator-friendly graphic form.

Sensor nodes shall transfer necessary information for understanding the network topology to the server on a periodic basis or by administrator's command. The information with the following contents is composed into database:

- The H/W-related table is defined to manage the H/W information and state of sensor nodes in the server.
- The information about sensor network topology is saved and managed in the server.
- The sensor network supporting the mobility of sensor nodes and sink nodes shall manage changes in network topology by movement of the nodes.
- The sensor nodes shall save information about composition management, and upon request from the server, deliver it to the server via gateway.
- The sensor node and gateway shall save their ID, and the server synthetically saves and manages the IDs of sensor nodes and gateway. These IDs should be provided in an administrator-friendly form (alias). For this, network management system shall provide mapping function between ID and alias.
- The server shall manage the location information of sensor nodes. Such location information may be entered by the administrator or automatically reported to the server by the location sensing function. The location information should be saved and managed in connection with ID of the sensor node of the server. Cloud computing is one such emerging paradigm which makes use of contemporary virtual machine technology.

Message switching to give time synchronization command from the sensor node at an upper level to that at a lower level should be provided. A list of neighbor sensor nodes should be obtained. The obtained list of sensor nodes contains Type (parent, child and neighbor nodes) information displaying the relation with each network as well as address of each node and information about RSSI. When obtaining list information, the maximum desired number of units should be clarified certainly.

Information about the sensor unit mounted on the sensor node shall be imported. The function of importing the condition of the sensing node and the sensing information from the network administrator should be provided.

Information about the sensor unit mounted on the sensor node is required to be set. Sensor sampling for the precision of energy management and sensing is also required to be set.

Primitive messages for information exchange between the server and the mobile terminal are defined. Each message is divided into Get command set to get information from the server and Set command set to set a certain value in the server. Each primitive message is as follows:

- Get Sensor Network Lists (Request, Response, Indication, Confirm): Message to get a list of sensor networks accessible from the server
- Get Node Lists (Request, Response, Indication, Confirm): Message to get sensor nodes present in certain sensor network
- Get Node Info (Request, Response, Indication, Confirm): Message to get information about a certain sensor node
- Get Node Data (Request, Response, Indication, Confirm): Message to receive data about a certain sensor node
- Get G/W Info (Request, Response, Indication, Confirm): Message to get information of the sensor gateway which manages certain sensor network
- Set Data for a Sensor Node (Request, Response, Indication, Confirm): Message to set and control information about a certain sensor node
- Set G/W Info (Request, Response, Indication, Confirm): Message to set and control information about a certain sensor gateway
- Status Message: Message to check the progress of the requested job

Communication between home network management server and embedded server is done by utilizing such primitive messages. Each message is properly used according to the job to be performed. Necessary procedure for performing the job is as shown in Fig. 1.

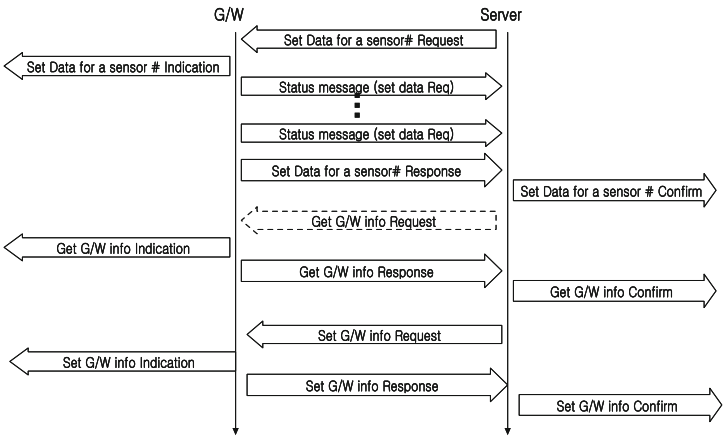


Fig. 1. USN message from the server to the gateway.

2.2 U-learning System Model

U-learning involves creating learning activities, tasks, projects, and resources that encourage students to discover learning themselves without consciously realizing they are learning, so that they learn automatically and independently. U-learning is a learning paradigm which takes place in a ubiquitous computing environment that enables learning the right thing at the right place and time in the right way. It is an expansion of previous learning paradigms as we move from conventional learning to electronic learning to mobile learning and now to ubiquitous learning (thus, the meaning of "e" is not just limited to "electronic" but expands to "everywhere," "extending", "enhancing," and "enabling." u-Learning enables us to change our current learning processes to be more efficient and more effective. If done right, u-learning becomes a critical force to improve the performance of our workforce and our organization as a whole [8]. The starting point of connectivism is the individual learner. Personal knowledge is comprised of a network, which feeds into organizations and institutions, which in turn feed back into the network, and then continue to provide learning to individual.

The following are the factors of the implementation to the u-learning model (see Fig. 2).

(i) Study contents management system and bi-directional contents.
(ii) U-learning contents for secondary study through the development of practically experiencing type contents based upon touch screen on tablet terminal and intuitional GUI
(iii) The actual feeling style English contents is interchangeable with the digital textbook and other studying management system, supporting various OS considering OSMU (One Source Multi-Use) and various device (dedicated terminal, common tablet, common notebook)
(iv) Utilization of practically experiencing type contents interchangeable to existing contents and study management solution
(v) Applicability in the various OS environments (Windows, Linux, Mac OS, etc.) since it is developed into open platform using open software

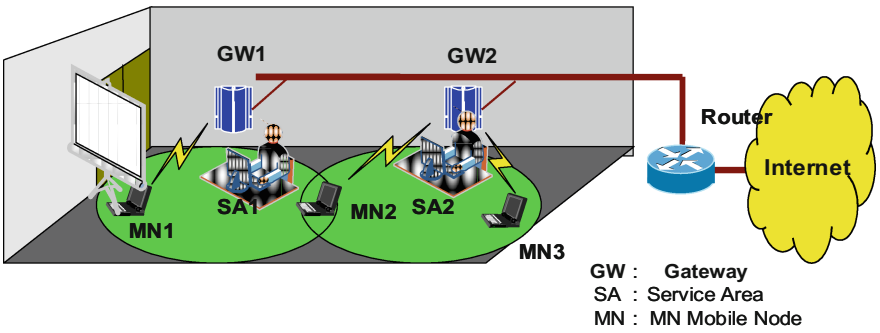


Fig. 2. USN system using u-learning.

(vi) Interchanging standard of contents with existing e-Learning standard
(vii) Operating in various equipment such as mobile equipment through optimization process

Existing ubiquitous network systems are mostly intended to gather and manage the environment information using the variant devices. However, true ubiquitous network should be capable of gathering and managing all environment-related room and outdoor information actually. While room information has less movement and focuses on data service compared to outdoor.

3 Conclusion

Ubiquitous network provides users with a variety of information services. Specially, kind and quality of the services can be abundant by utilizing various sensor data. However, currently, ubiquitous network itself limits available range of the services by concentrating upon multimedia service. Specially, it has achieved huge performance in controlling the appliances by focusing on the communication network but obtained little performance in controlling the appliances.

For this, protocol to manage network in ubiquitous sensor network was defined and implemented for wireless home network in this paper. Although network management system in USN related with this paper is being progressed for the purpose of independent projects, protocol interface and message system have not been clearly defined yet such as general purpose network or extension into heterogeneous kinds, communication support etc.

Based on the mentioned, required management items, PI (Protocol Interface) for efficient and systematic communication between networks was defined and the subsequent message was described.

Thanks to the information technology, the e-learning technology is evolving gradually to a form of the u-learning while the ubiquitous environment is accelerating. It is expected that the e-learning supportive platform which can be used for studying using the terminal (cellular phone, PDA, laptop computer, PMP, etc.) at anytime, anywhere will be appeared in the market.

References

1. Khan, B.H.: A framework for web-based learning. Educational Technology Publications, Englewood Cliffs (2000)
2. Liaw, S.-S., Huang, H.-M., Chen, G.-D.: Surveying instructor and learner attitudes toward E-learning. Comput. Educ. **49**, 1066–1080 (2007)
3. Wang, Y.-S., Wang, H.-Y., Shee, D.Y.: Measuring e-learning systems success in an organizational context: Scale development and validation. Comput. Hum. Behav. **23**, 1792–1808 (2007)
4. Chikh, A., Berkani, L.: Communities of practice of e-learning, an innovative learning space for E-learning actors. Procedia Soc. Behav. Sci. **2**, 5022–5027 (2010)

5. Sultan, N.: Cloud computing for education: a new dawn? Int. J. Inf. Manage. **30**, 109–116 (2010)
6. Arshad, J., Townend, P., Jie, X.: A novel intrusion severity analysis approach for Clouds. Future Gener. Comput. Syst. (2011). doi:10.1016/j.future.2011.08.009
7. Loganayagi, B., Sujatha, S.: Enhanced cloud security by combining virtualization and policy monitoring techniques. Procedia Eng. **30**, 654–661 (2012)
8. Yahya, S., Ahmad, E.A., Abd Jalil, K.: The definition and characteristics of ubiquitous learning: a discussion. Int. J. Educ. Dev. Using Inf. Commun. Technol. **6**(1), 117–127 (2010)

A Quality Model for IoT Service

Mi Kim[1], Jin Ho Park[2], and Nam Yong Lee[2(✉)]

[1] Department of Computer Science, Graduate School, Seoul, Korea
pytwoori@gmail.com
[2] School of Software, Soongsil University, Seoul, Korea
{j.park,nylee}@ssu.ac.kr

Abstract. In this paper, We focuses on suggestion to Quality model of IoT Services which are popularly used in the characteristics defined in the quality model of IoT Services. In order to achieve this purpose, we propose Quality model for IoT Services. These technologies involves utilization and mobility in addition to quality characteristics in existing software, application of ISO 9126 is not perfect when evaluating a Quality model. This paper proposes a security set out in ISO25000 quality factors and assessment of the existing traditional software application of ISO 9126 quality model. We suggested new quality model for IoT Services by quality attribute in ISO 9126. We validated that the proposed model can be realized it was applied to evaluate the 4 elements and added security in Metrics. The quality model for IoT Services using the IS-QM proposed in this paper it can be measured relatively accurately.

Keywords: QoS model · IoT Services · Internet of Things · Quality attribute · Software Quality

1 Introduction

Internet of Things (IoT) is the computing environment to provide valuable services by interacting with various IoT Services, where diverse devices are connected within the existing internet infrastructure and through intelligent social applications. A Quality model for Software applications acts as a framework for the evaluation of attributes of an applications that contribute to the quality model. It is important every relevant software applications quality characteristic is specified and evaluated whenever possible using validated or widely accepted metrics. It is necessity to customize a quality model to identify acceptance criteria and evaluate a particular application domain; IoT applications. A natural consequence of the trends would be to manage the quality of IoT Services. However, measuring the quality of IoT Services is considerably different from measuring the quality of conventional software systems. Because IoT Services is a complex fusion of a variety of technologies such as wireless network, embedded, sensor and connectivity. For instance, ISO 9126 which is a representative quality model is mainly for conventional software, not addressing the IoT characteristics. That is the quality attributes and metrics in ISO 9126 would largely inappropriate in measuring the quality of IoT Services. We suggested new quality model for IoT Services by quality attribute in ISO 9126. We validated that the proposed model can be realized it

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_77

Fig. 1. IoT services

was applied to evaluate the 4 elements and added security in Metrics. We proposes a security set out in ISO25000 quality factors and assessment of the existing traditional software application of ISO 9126 quality model. In this paper, we intend to propose a security set out in ISO2500 quality model to identify characteristics of quality factors which are quality attributes, criteria and metrics evaluate a particular IoT Services (Fig. 1).

2 Related Works

A Quality model has been studied by many researchers in conventional software. ISO/IEC 9126 is not perfect when quality of IoT services. The Quality model can analyze IoT contexts and employ methods to compute the value of QoS which acts as a metric for service evaluation and selection. But most previous work focused on the RFID network protocols, middleware, devices reliability, safety and cost, etc. Evan Welbourne deployed these applications in the RFID Ecosystem and conducted a four-week user study to measure trends in adoption and utilization of the tools and applications as well as users' qualitative reactions [1].

In [2], the author constructed another AHP QoS model based on IoT global etc. were considered but users' feedback like user preferences was often ignored.

Simple Performance Analysis of Multiple Access RFID Networks Based on the Binary Tree [3] This simplifies tremendously the derivation of the analysis, still preserving its validity in an average sense. scalable RFID network simulator is developed with OPNET 11.0 and implements the binary tree collision arbitration protocol. The results obtained with simulator are very complexity. Performance evaluation is conducted for tag read latency and read efficiency as a function of link parameters defined in ISO/IEC 18000-6 Type C standard [4] But this ISO9126 is difference for application quality evaluation. Reference [5] first proposed a QoS model of grey decision-making from the view of IoT Global Infrastructure and built an adaptive service framework. In [6], a quality evaluation technique of RFID middleware according to ISO/IEC 9126 standard and EPC global middleware quality factors was built by simple AHP. It really provided an idea of simplifying complicated evaluation factors in IOT. ISO 9126 is an international standard for the evaluation of product quality [7]. This standard provides three aspects for evaluating software products; internal quality, external quality and quality in use. And, there are sixteen characteristics for three types of qualities. However, this standard focuses on evaluating quality of conventional products.

Hence, it is required that the standard is customized and extended to evaluate the quality of IoT Services.

3 Characteristics of IoT Services

IoT Services reveal non-conventional features which are not typically presented in conventional software systems. it is as in the following.

3.1 Participation of Hardware Devices

In IoT, things are required to be actively participating in various activities, exchanging information and making intelligent decision. IoT devices are equipped with various intelligent devices to ensure safety, security and high performance of IoT devices. Mobile devices are equipped with small buffers but have to deal with IoT services that generate a huge volume of data Hence, an IoT Services consists of two types of elements; typical software components and hardware devices/components. In the design and implementation of IoT applications, the presence of hardware devices should be considered.

3.2 Collaboration Model with IoT Devices

A software application typically consists of multiple *business processes*, also called *workflow*, *collaboration*, *orchestration*. In IoT Services, collaborations could include the hardware functionality of IoT devices. For example, the *Drone-based Parcel Delivery System* by *Amazon* includes the parcel pick-up service and flying service of *Drone* devices. The collaborations of an IoT Services should consider the hardware functionality as well as conventional software collaboration.

3.3 Mobility and Connectivity

It refers to the application capability for device mobility the correctness of the information processing in different stages of a process. This characteristic is merely used in the web application and particulary in IoT services. Connectivity characteristics refers to the ease of user's quick and efficient connect to information of IoT services. The mobility could result in various faults; loss of control due to out-of-sight movement or flight, unstable wireless network connection when leaving the current network zone, and physical collision while moving.

3.4 Monitoring for IoT Devices

Remote monitoring for smart devices attached to networks can support multiple functions including Automatic Meter Reading, gas heating and context-awareness.

Through smart devices, carriers can provide more real-time and precise services. It has a characterized that connectivity to the objects for control. An IoT device can controls flow of the each data through a collaboration device.

3.5 Limited Resources

The resource types of IoT devices can be battery, network communication facility, memory, and computation power. For example, flight Drone has a limited battery and memory, leaving the current network zone. IoT devices suffer from a limited battery lifetime and dominates energy consumption. Energy Efficiency can be increased by wisely adjusting transmission power. Hence, there is a need for solutions that will limit the energy consumptions of such IoT devices such as wearable device, smart watch.

3.6 Security and Certification

For the security and collaboration of IoT device become a participation of device that they need the process can be exact service of device. Among collaboration of device for the security from intermediate extortion can be need to authentication procedure of precise device. In the cause of prevent intermediate extortion can be need to designed consideration of the authentication algorithm. Figure 2 shows that the characteristics of IoT Services and each criterion is defined in this section.

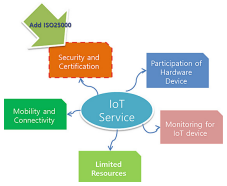


Fig. 2. Characteristics of IoT services

4 Quality Attributes and Metrics

The characteristics defined in the previous section become the basis for deriving the quality model of IoT Services. That is, we define quality attributes for evaluating IoT Services by considering the impacts of the identified characteristics on the quality of IoT Services.

4.1 Quality Attributes for Evaluating IoT Services

Each characteristics of IoT Services has a set of associated quality factors. Figure 3 shows how the quality attributes of the evaluating IoT services affect the selection of

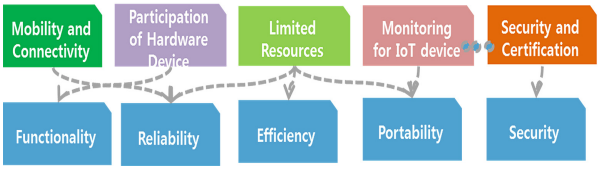


Fig. 3. Mapping characteristics to quality attributes

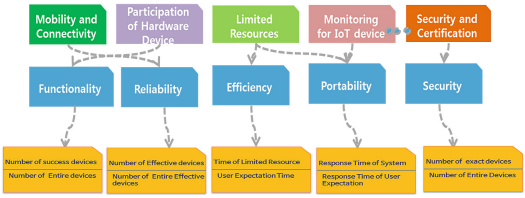


Fig. 4. Main Factor characteristics to quality metrics

quality factors. Figure 4 shows the factor characteristics to quality metrics, and each criterion is defined in this section.

4.1.1 Metric Suite for the Quality Model

In this section we provide a set of representative metrics for the criteria like all quality models, the set of proposed metrics here may be extended or tailored for each organization. Hence, the proposed set of attributes should serve as the foundation and a starting point for tailoring and enhancement.

4.1.2 Metrics for Functionality

Metric for Functionality: this metric measures the degree of functional connection to IoT service, as follows. X = A/B, where A is the number of correctly accessibility related to functionality of characteristics for IoT services is applicable to, and B is the total number of functional accessibility in the IoT domain. The Range of X is 0..1, and the value 1 is the best since the functionality of such a application can be accessed to all functional items in the IoT domain. The higher the metric value has wider range of accessibility.

4.1.3 Metrics for Reliability

Metric for Reliability: The metric measures how the without seamless each IoT application connection as follows: X = A/B, where A is the number of input items which check for valid data, and B is the number of total input items which could check for valid data.

The Range of X is 0..1, and the value 1 is the best. Although practically data will have a value that is high due to validation data.

4.1.4 Metrics for Efficiency

Metric for efficiency: The metric measures the efficiency appropriate time and resource behavior. X = A/B, where A is the time of limited battery utilization, and B is the time of user requirement time and resource. The Range of X is 0..1, and the value 1 is the best, The time and resource utilization describes for instance processing times and throughput rate, while resource behavior means the amount of resource used and the duration of use.

4.1.5 Metrics for Portability

Metric for conformance: The metric measures the degree of change to relevant environments. X = A/B, where A is the number of system response request which can be customized, and B is the user expected response requirements number can be changed relevant environment system as a sensory data creative value in the IoT Services. The Range of X is 0..1, and the value 1 is the best, response request number for heterogeneous system such as a sensory data through IoT Services is higher for system utilization.

4.1.6 Metrics for Security

Metric for conformance: The metric measures the degree of change to relevant environments. X = A/B, where A is the number of Exact device which can be customized, and B is the entire device can be accessed relevant environment system as a sensory data creative value in the IoT services. The Range of X is 0..1, and the value 1 is the best, the number of exact device is close to precision to system such as a exact data through IoT Services is higher for system accessibility.

4.1.7 Sum of IoT Service Metrics

Using the proposed a set of factors and criteria, one can compute the overall quality of a IoT Service as follows. Let Fi be the ith factor defined in IS-QM, and therefore the range of i is 1..5. Let Cij be the jth criterion of the Fi, and Fi has 2 to 5 criteria as shown in Fig. 4. For example, C2, 3 will be the customizability criterion for the reliability factor. Then the composite value, Q Fi for the factor Fi is defined as:

$\mathrm{Q\,F}_i = \frac{\sum_{j=1}^{n} Cij}{n}$ where n is the number of criteria in $\mathrm{F_i}$.

Depending on the domain and purpose, often different weight on $\mathrm{F_i}$ and let (W1 + W2 + W3 + W4 + W5) be equal to 1. Then, the overall quality of a functions can be computed as follows, and the range of overall quality will be 0..1, which is same as a single percentage representation.

$$IS - QM = \sum_{i=1}^{5} wi * qi$$

In order to apply the proposed IS-QM effectively, a quality evaluation process, instruction and a set of templates should also be defined. As shown in Table 1, Weight for the quality attributes such as high(0.3), medium(0.2) and low(0.1) would be represented. The IS-QM model has value 1, the number 1 is evaluated with high quality.

Table 1. Weight for the Quality attributes

Quality attributes	Weight
Functionality (Func)	Medium (0.2)
Reliablity (REL)	High (0. 3)
Efficiency (Eff)	Medium(0.2)
Portability (Por)	Low (0. 1)
Security (Sec)	Medium(0.2)
SUM	1

5 Conclusion

In this paper, we showed identified main characteristics of IoT services, and derive a practical quality model for IoT services, IS-QM. The IoT Services is a complex fusion of a variety of technologies such as wireless network, embedded, sensor and connectivity. Because these technologies involves utilization and mobility in addition to quality characteristics in existing software, application of ISO 9126 is not perfect when evaluating IoT services. We suggested new quality model for IoT Services by quality attribute in ISO 9126 and ISO25000. We defined five quality factors, criteria, and five metrics (Func, Rel, Eff, Por, Sec) that can be effectively used to derive the overall quality of IoT Services. The effectiveness of the quality model for evaluating IoT Services through quality attribute-based metrics were validated. the quality model for IoT services using the IS-QM proposed in this paper it can be measured relatively accurately.

References

1. Welbourne, E., Battle, L., Cole, G., Gould, K., Rector, K., Raymer, S., et al.: Building the internet of things using RFID: the RFID ecosystem experience. 2009. IEEE Internet Comput. **13**, 48–55 (2009). On Modern Computing (JVA 2006), pp. 163–168 (2006)
2. Shaoshuai, F., W, S., Nan, W., Yan, L.: MODM based evaluation model of service quality in the internet of things. Procedia Environ. Sci. **11**, 63–69 (2011)
3. Ferrari, G., Cappelletti, F., Raheli, R.: Simple performance analysis of multiple access RFID networks based on the binary tree protocol. Int. J. Sensor Netw. **4**, 194–208 (2008)
4. Ko, C., Roy, S., Smith, J.R., Lee, H.W., Cho, C.H.: RFID MAC performance evaluation based on ISO/IEC 18000-6 type C. IEEE Commun. Lett. **12**(6), 426–428 (2008)
5. Liu, J., Tong, W.: Adaptive service framework based on grey decision-making in the internet of things. In: 2010 6th International Conference on Wireless Communications, Networking and Mobile Computing, WiCOM (2010)
6. Oh, G., Kim, D.Y., Kim, S.I., Rhew, S.Y.: A quality evaluation technique of RFID middleware in ubiquitous computing. In: 2006 International Conference on Hybrid Information Technology, vol. 2, pp. 730–735 (2006)
7. Software Engineering – Product Quality – Part 1: Quality Model. ISO/IEC 9126, June 2001

Author Biographies

Mi Kim received the Master Degrees from the University of Honam, in 2000 and 2002, PhD at the department of Computer Science at Soongsil University in Seoul, Korea. She had worked in LG Electronics Inc in 1996 and 1999 and now she is a Project Manager and Programmer at the IT Solution Cor. Hers research interests include software engineering, IoT Applications, Software Quality.

Jin-Ho Park received his bachelor's degree of Software Engineering in Soonsil University, Seoul (1998), and master's degree (2001), doctor's degree of Computer Science in Soongsil University, Seoul (2011). Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on Software Engineering, SW Safety/QA/Testing, SW Convergence/Power, IoT service.

Nam-Yong received his bachelor's degree of Computer Science in Soonsil University, Seoul and master's degree of MIS in Graduated Korea University, Seoul. Doctor's degree of MIS in Mississippi State University, USA. Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on System Engineering, Software Engineering, E-Commerce System, MIS, etc.

An Empirical Study of Risk Factors for the Development Methodology for Small-Size IT Projects

Joon Ho Park[1], Nam Yong Lee[2], and Jin Ho Park[2(✉)]

[1] Department of Computer Science, Graduate School, Seoul, Korea
108.joonho.park@gmail.com
[2] School of Software, Soongsil University, Seoul, Korea
{nylee,j.park}@ssu.ac.kr

Abstract. The existing theoretical methodology or development methodology for small-size IT project when applied to the creation of project development deliverables, and many people are spending time. Because of enterprise projects to proceed in accordance with the procedures of the methodology, the lack of induction person or projects develop document written with experienced personnel is insufficient. In addition, when applied to development methodology of enterprise submitted a large amount of project development deliverables, and development of personnel development should be put into a lot of time on unnecessary paperwork to consume the result is displayed. This in the Dissertation, theoretical development methodologies and development methodology of domestic enterprises to use seven mandatory development deliverables and options development deliverables by analyzing the derives. In addition, client and expert of interviews, the developers from the survey will provide current problem and direction. Through which small businesses can easily leverage the development of the project development methodology, and short-term projects, while small and medium enterprises development deliverables written to create an efficient procedure is defined. Each stage through which small businesses can derive an important development deliverables, -based project development methodologies suitable for small and medium enterprises can take advantage of the methodology.

Keywords: Project methodology · Development methodology · Software engineering

1 Introduction

Development methodology for small-size IT project is standardization of techniques and utilities, which are needed to manage the process of performing projects for system integration (SI) or development. A development methodology for small-size IT project consists of multiple components such as task procedures, task methodology, project development documents, management method, and development tools etc. A development methodology for small-size IT project is composed of procedural steps, methods, standard templates, quality control and configuration management, and a

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_78

project is conducted with the human resources who are able to follow a development methodology for small-size IT project. development methodologies for big-size IT projects or medium-sized IT projects are appropriate for large-scale IT projects that are usually conducted by big companies. However, the development project that is conducted by a small-size company is supposed to be completed with low budget, short period and the minimized human resources. Even for the same development methodology for small-size IT projects, a development methodology for small-size IT project can be differed in task procedure and task method etc. according to business scale and development processes. If previous development methodologies were uniformly applied for every project without considering business scale and development processes, then human resources and time would be wasted for delivering project development products (s, documents, and etc.). Because small-size companies do not have constraint human resources who can lead project by following the procedures of previous development methodologies, or who have experience of the deliverables guide for s, documents, and etc. And also previous development methodologies require the large number of s, documents, and etc… Moreover, small-size companies adopt previous development methodologies or theoretical development methodologies for small-size IT projects without modifications and considerations of their project environment, but those kinds of development methodologies for small-size IT projects are not effective or productive, and do not deduces the development failures. Therefore, a new development methodology for small-size IT project is needed that considers the status of small-size IT project environments for efficiency and productivity.

This paper consists of seven sections. Section 2 describes the background and previous work. Section 3 presents the customer interviews. Section 4 presents the developer survey results. Section 5 presents the implications for customers and developers. Section 6 presents a development methodology for small-size IT projects. Section 7 will be described regarding how to avoid the risk factors.

2 Previous Work and Analysis

In general, most of domestic solution development methodologies for small-size IT project have used the previous methodology that is adopted from foreign countries and revised according to big-size company's circumstances and needs. Since previous methodologies of foreign countries are established by the accumulated experiences and know-hows in long term, adaptation of the methodologies should be considered and could be modified based on the enterprise's circumstances and needs. The representative solution development methodologies of foreign countries are METHOD/1 (Andersen Consulting), NAVIGATOR (Ernst & Young), IEM (James Martin) and 4FRONT (Deloitte), and solution development methodologies devised for domestic circumstances are Innovator (Samsung SDS), POS-IEM (Posdata), SLC (LE-EDS) and HIST4FRONT (Hanjin Information Systems & Telecommunications) etc.

In this paper, in order to analyze solution development methodologies, we categorize them in according to the number of employees. We think that the number of employees means a project size and budget that the company is able to participate in.

Therefore, the previous solution development methodologies can be categorized into three folds.

1. Large size companies- level A: more than 2,500 employees
2. Medium size companies- level B: 300 to 2,500 employees
3. Medium and small size companies-level C: Under 300 employees

In this paper, through analyzing five solution development methodologies of level A (eGovFrame, CBD, LG CNS, Samsung SDS, and Marmi IV) and level B (Dongbu CNI and Kolon Benit), we induce the issues for making solutions, documents, and etc. and procedures when we apply the above solution development methodologies to short term solution projects. Samsung SDS of Innovator development methodology consists of 4 steps totally and 13 procedures, 23 tasks and 27 solutions, documents, and etc. are induced. For completing development project by the Innovator, the total service period becomes 120.6 days when the duration time of the solution development project for small-size IT project persists six months. This indicates that each solutions, documents, and etc. is supposed to be completed for 4.47 days arithmetically to complete the 27 solutions, documents. Since a small-size company determines the human resources and project time according to project budgets, the small-size company cannot assure enough time to increase the quality of solutions, documents, and etc. and to conduct the project simultaneously. And also the numbers of employees who have experience of various tasks, deliverables and solutions are not enough.

We analyze seven big 7 development methodologies that are commercially and theoretically recognized as legacy development methodology. We analyze each development methodology's development stages, procedures and solution and deliverables. And we adjust each development methodology's development stages and procedures for comparison, since they have slightly different stage name and procedure name for same task and activities (Fig. 1).

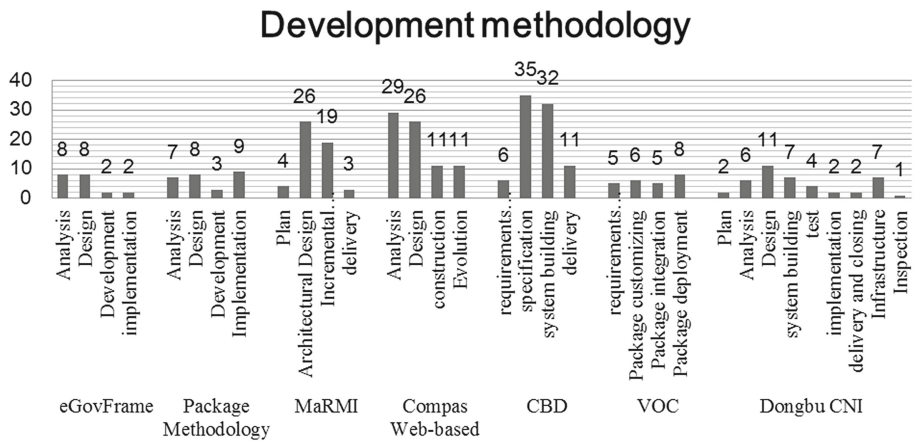


Fig. 1. Development methodology results

3 Customer Interviews

Customers do not consider the project deliverables after the project completion. However, the deliverables are important for managing the completed operating system. Finally, the documents for systems operation is important, it is needed to write the operation guide which must be recorded a large amount of information. We found the problem of the project procedure through the customer interviews. All the interviewed customer response the same answer that the developed product features are different from those they intended. Finally, due to the unspecified development and the designed deliverables, the customer requirements in the analysis phase results in the development of the un-wanted products (Table 1).

Table 1. Customer interview results

Division	
Important task	Analysis is important
After completion of the project important development deliverables	Operator's guide, Operator's manual
Requirements reflect the time	From the development task

4 Developer Survey Results

Developers Online Survey results can make the following conclusions. The conclusion that can reflect customer requirements for our customers and reflect designs were drawn up differently. Education is currently developing methodologies and solutions have been proposed to proceed differently than a lot of good comments on the analysis and design phase.

5 Implication

Developers surveys and customer interviews through the development of methodology, customers were deemed important to the analysis and design phase. However, in the case of the developers of the project, or no experience in the analysis and design of the business process steps can be seen that the lack of understanding.

6 Risk Factors Improvement of Small-Size IT Projects

Development deliverables presented in this paper are a total of 6 activities in three phases, nine operations, it was presented to the 13 development outcomes. This paper presents the development work and deliverables for each task in the work and activities to be performed in each of the activities. Table 2 proposed development deliverables can be used to tailoring to suit the characteristics of the project.

Table 2. Development deliverables list

Task	Activity	Document
Analysis (Requirements) and design	Requirement define	System analysis
		Interview results
		Requirements analysis
	Architecture design	Architecture design
	Database design	ERD
		Entity design/List
		Table design/List
	Test	Test plan
Development	Development	Source code
	Test	Unit test scenario/Results
		Integration test scenario/Results
Implementation (operation)	System deployment	Education plan
		Pilot operating plan and journal

7 Conclusion

For IT projects for service s or system integration, there are many kinds of IT service development methodologies that are developed and customized by each global company that has sufficient human resource and experts and sufficient budget. Thus these kinds of IT service development methodologies are appropriate to small-size project.

In this paper, we present development methodology for small-size IT project that is appropriate to small-size companies that have constraint of human resource and budget. St first step, we analyze previous IT service development methodologies that are used for big-size companies and previous theoretical IT service development methodologies. And we have interviews with many experts and customers who present and propose their opinions for mandatory and optional deliverables for each stage. From the analysis and the interviews, we drive mandatory deliverables and optional deliverables for each development stage.

Those project procedures and stages are appropriate for short-term projects and through the three stages that are newly proposed in this paper are integrated with the analysis and design procedure. In the analysis stage, education task by the introduction and education of the solution to the customer, so that customers are aware of the specific requirements in order to understand the s and products of the short-term projects. From detailed requirements analysis, the frequency of project plan and development changes can be reduced, that occur in the development phase and customers' requirements can be satisfied. By following proposed development methodology for small-size project, the high quality of short-term project could surely be guaranteed.

We have a plan to design and development project management tools and standards deliverables format.

References

1. Choi, M.J.: A study on the application effect and improvement of agile methodology for project management. Master's thesis, Hanyang University, Graduate School of Engineering (2011)
2. Park, S.J.: A study on the tailoring of object-oriented software development methodology foe small/medium scaled projects. Master's thesis, Sogang University Graduate school of information & Technology, Seoul, Korea (2000)
3. Lim, Y.T.: Development of Engineering Methodology Using Process Extraction Model for Small to Medium Size Organizations. Ph.D. thesis, The graduate school of ajou university, Seoul, Korea, (2011)
4. Lee, J.Y., Park, S.H.: The IT services industry globally competitive Strengthening. Research Paper, Korea IT Service Industry Association, Seoul, Korea (2010)
5. Kim, D.H.: The Audit Method for Improving The project Quality by Applying the agile Methodology. Master's thesis, Konkuk University Graduate school of information & Technology, Seoul, Korea (2011)
6. Robey, D., Welke, R., Turk, D.: Traditional, iterative, and component-based development: a social analysis of software development paradigms. Inf. Technol. Manage. **2**, 53–70 (2001)

Author Biographies

Joon Ho Park received his bachelor's degree of Mobile Communication Engineering in Dongyang University, Gyeongbuk (2003). He is Master Degree of Information Science in Graduated Korea National Open University, Seoul (2015). He is studying Ph.D. software engineering in Graduated Soongsil University, Seoul. From 2016 until now, from the work in the Bigdata analysis, CF Algorithm, Development Methodology Process, the main areas of interest software engineering, development methodologies, Bigdata analysis, etc.

Nam-Yong Lee received his bachelor's degree of Computer Science in Soonsil University, Seoul and master's degree of MIS in Graduated Korea University, Seoul. Doctor's degree of MIS in Mississippi State University, USA. Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on System Engineering, Software Engineering, E-Commerce System, MIS, etc.

Jin-Ho Park received his bachelor's degree of Software Engineering in Soonsil University, Seoul (1998). and master's degree (2001), doctor's degree of Computer Science in Soongsil University, Seoul (2011). Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on Software Engineering, SW Safety/QA/Testing, SW Convergence/Power, IoT, National Defense ISR, IT Service, IT Technical Commercialization and Start-up, etc.

Nack-Based Broadcast Mechanism for Isochronous Audio Stream Transmission Using Bluetooth Low Energy

Jaeho Lee(✉)

Department of Information and Communications Engineering,
Seowon University, Cheongju, Korea
izeho@seowon.ac.kr

Abstract. Bluetooth Low Energy which has been developed and deployed into the real home environment provided high energy efficiency where only low traffic environment was needed. However, the current version of this technology couldn't support audio broadcast stream transmission because this environment requires classified isochronous data channel. For addressing this issue, this paper propose a new broadcast mechanism under the assumption of the next generation technology of Bluetooth Low Energy, using Nack mechanism combined with energy detection mechanism, in order to support seamless broadcast transmission of audio data stream. For further study, this paper designed energy detection for Nack reception without awareness of the number of slave nodes.

Keywords: Networks · Hybrid routing · Routing · Traffic-aware

1 Introduction

Todays, Bluetooth technology has been widely deployed to not only the smart home area including smart office environment but also various wireless personal area networks. We can see anywhere people are using various Bluetooth-based mobile devices such as wireless headphone including microphone to be connected to handheld devices or smart watch to be also connected to smart phone or laptop. As a result, Bluetooth technology was designed for seamless data transmission especially to audio stream service.

By the way, Bluetooth SIG (Special Interest Group) published a new technology named as Bluetooth Low Energy, which can provide lower energy consumption than the previous Bluetooth technology. And many people forecast that Bluetooth Low Energy technology will occupy the market of the previous Bluetooth devices. However, Bluetooth Low Energy technology was basically designed for connectionless communication or short-term event-driven communication use-cases.

In smart home area, there are many needs of short range broadcast data transmission such as TV to individual headphones of multiple users and teleconference devices to individual headphones and microphones of multiple users. In this case, Bluetooth Low Energy technology couldn't provide guaranteed seamless audio stream because the master devices couldn't know how many slaves existed and also couldn't know which slave didn't receive the current packet.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_79

For this issue, this paper provide a Nack-based broadcast mechanism based on energy detection method for all unknown slaves.

2 Protocol Description

2.1 Broadcast Transmission Environment

Audio stream has the characteristics of seamless occurrence, and this doesn't require high speed or high congestion control but just only require no delay time for the transmission. However, the current version of Bluetooth Low Energy has significant disadvantage of protocol overhead to be completely established, as shown in Fig. 1. Hence this paper employed isochronous channel mechanism which will be described in the next version of Bluetooth Low Energy standardization, as shown in Fig. 2.

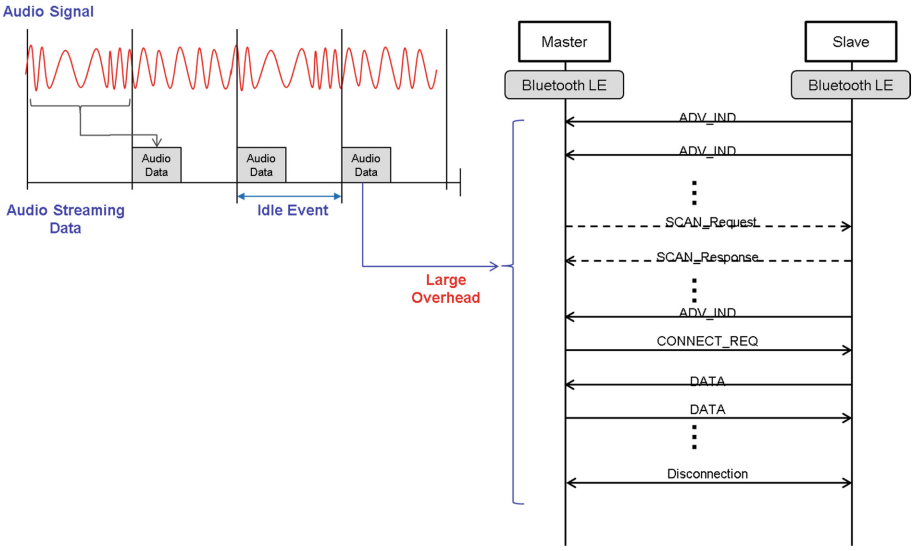


Fig. 1. The difference of zone shape

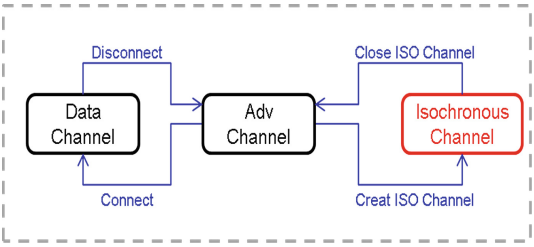


Fig. 2. The state machine of isochronous channel

2.2 Motivation

Currently, Bluetooth SIG has developing the isochronous broadcast mechanism approaching the N-repetition retransmission from the master device because it couldn't know how many slave devices existed. However, isochronous broadcast is used for not only public broadcast use-case but also multi-user HQ audio scenarios such as TV to multiple headphones which also employed Bluetooth Low Energy technology.

In the case of that more than 2 users hear HQ audio via isochronous broadcast, mobile device, e.g., phone, watch. Etc., is used for isochronous source role, and it's clear that the energy efficiency of master device has to be considered. And above source device could be more energy-constraint in future. For any other reason, we would like to see the energy efficiency of master side. Hence, this paper was motivated to think about Nack-based isochronous broadcast for master device of Bluetooth Low Energy.

2.3 Energy Detection Method

As we know, N-times retransmission method is an appropriate solution where master couldn't know the number of slave devices. Master device couldn't guarantee that all slave devices successfully receive all data transmission. However, if the quality of the radio channel condition is clear, this mechanism can lead to overhead incurred from N-times retransmissions, even though this is very robust but master would consume more energy unnecessarily. Hence this paper propose Nack-based broadcast mechanism, as shown in Figs. 3, 4, and 5.

In this protocol, only one slot for Nack reception on master device is needed in each sub-isochronous event. And, master device detects energy level which increased than the given threshold value, it assume some slave devices didn't receive data so they send Nack.

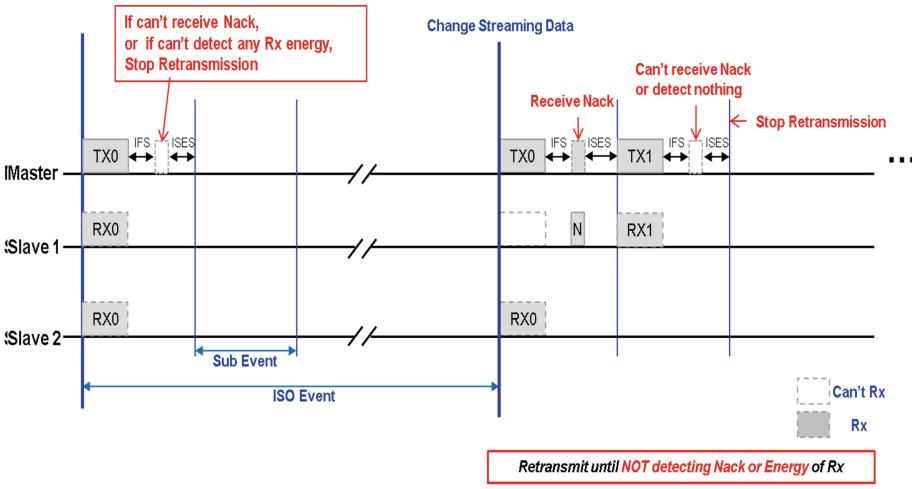


Fig. 3. The example of nack-based broadcast – single slave didn't receive data

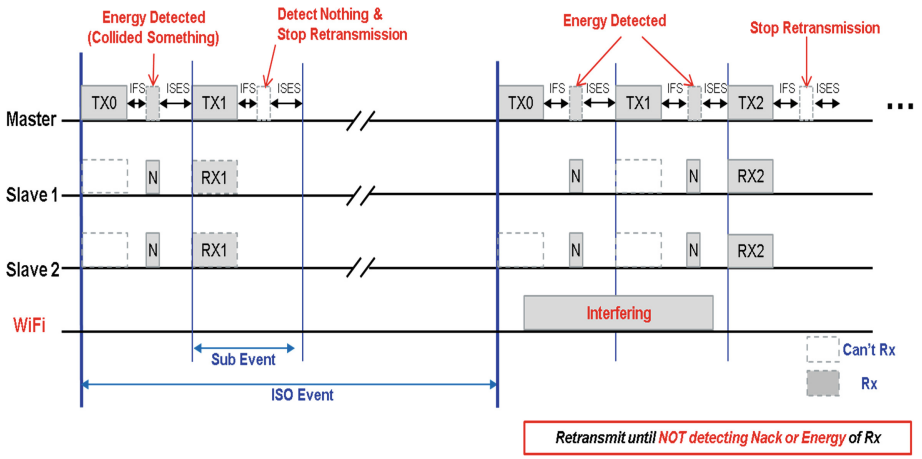


Fig. 4. The example of nack-based broadcast – external interference

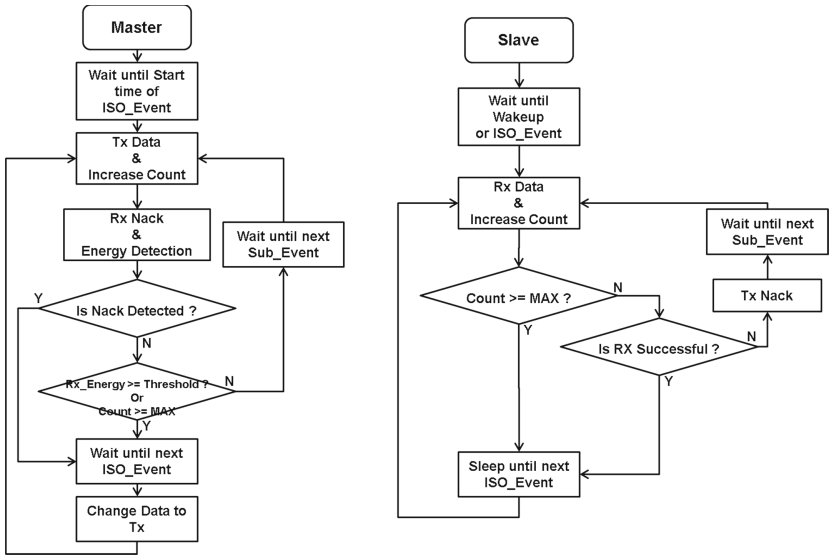


Fig. 5. The algorithm for nack-based broadcast mechanism

In this case, master device surely transmit again. On the other hand, if master device couldn't any energy level during this period, it will stop to transmit data because any slave device did not fail to receive data so any slave did not send Nack frame. On the slave device, it do nothing and enter to sleep to save its energy after successfully receives data from the master device, and will send Nack frame in the case of vice versa.

Moreover, we can assume that the event of Nack frame collisions can composed of two exceptional cases. First collided reason was derived from the collisions between multiple Nack frames sent from multiple slave devices. In this case, nevertheless of which Nack was collided, master device can assume that more than 2 slave devices couldn't receive data. Consequently master device just does retransmission without any concern.

Second case of that Nack frame collision is derived from external interference. However, this situation couldn't explain that a Nack was sent from any slave device or that external interference negatively influenced last data transmission. Hence, the master device just does retransmission. As a result, in case of master device detect any energy level after data transmission, it transmits again the same data without any technical doubt.

3 Conclusion

In this paper, proposed Nack-based mechanism can address unnecessary additional retransmission problem which can be incurred on previous N-retransmission method. With this mechanism, various needs for multicast or broadcast not only in short communication range such as home but also wide communication range such as public area will be improved via Bluetooth Low Energy with maintaining high energy efficiency. For further research, this energy detection-based Nack mechanism will be researched to be applied to Wi-Fi communication environment.

References

1. Bluetooth Core Specification 4.2, Bluetooth SIG, December 2014
2. Akyildiz, I., et al.: A survey on sensor networks. IEEE Commun. Mag. **40**(8), 102–114 (2002)
3. Ye, W., Heidemann, J.S., Estrin, D.: An energy-efficient MAC protocol for wireless sensor networks. In: Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2002), pp. 1567–1576, June 2002
4. Ye, W., Heidemann, J.S., Estrin, D.: Medium access control with coordinated adaptive sleeping for wireless sensor networks. IEEE/ACM Trans. Netw. **12**(3), 493–506 (2004)
5. Buettner, M., Yee, G.V., Anderson, E., Han, R.: X-MAC: a short preamble MAC protocol for duty-cycled wireless sensor networks. In: Proceedings of the 4th International Conference on Embedded Networked Sensor Systems (SENSYS 2006), pp. 307–320 (2006)
6. Yang, S.-H., Tseng, H.-W., Wu, E.H.-K., Chen, G.-H.: Utilization based duty cycle tuning MAC protocol for wireless sensor networks. In: IEEE GLOBECOM 2005 Proceedings, pp. 3258–3262 (2005)
7. Yoo, H., Shim, M., Kim, D.: Dynamic duty-cycle scheduling schemes for energy-harvesting wireless sensor networks. IEEE Commun. Lett. **6**(2), 202–204 (2012)
8. IEEE, IEEE Standard for Local and Metropolitan Area Networks, Part 15.4 (Low-Rate Wireless Personal Area Networks), September 2011
9. Zigbee Specification, Zigbee Alliance Inc., September 2012
10. https://Bluetooth.org

11. Alonso, G., Kranakis, E., Wattenhofer, R., Widmayer, P.: Probabilistic protocols for node discovery in ad-hoc, single broadcast channel networks. In: Proceedings Parallel and Distributed Processing Symposium, April 2003
12. Ding, G., Sahinoglu, Z., Orlik, P., Zhang, J.: Tree-based data broadcast in IEEE 802.15. 4 and ZigBee networks. IEEE Trans. Mob. Comput. **5**(11), 1561–1574 (2006)
13. Ghoshdastider, U., Viga, R., Kraft, M.: Experimental evaluation of a pairwise broadcast synchronization in a low-power cyber-physical system. In: 2015 IEEE Topical Conference on Wireless Sensors and Sensor Networks (WiSNet), January 2015

Empirical Study on IoT-Learning for the Rehabilitation Treatment of Chronic Low Back Pain Patients

Seul-Ah Shin[1], Ji-Soo Choi[2], Young-Jong Kim[3(✉)], Nam-Yong Lee[3], and Jin-Ho Park[3]

[1] Department of Computer Science, Graduate School, Soongsil University, Seoul, Korea
sashin@ssu.ac.kr
[2] Graduate School of Information Science, Soongsil University, Seoul, Korea
doddang2@ssu.ac.kr
[3] School of Software, Soongsil University, Seoul, Korea
opensys@gmail.com, {nylee, j.park}@ssu.ac.kr

Abstract. As the study of the IoT evolves, it is naturally used in everyday life. It is common situation that we can easily turn off the light as well as increase the temperature in the house by using a mobile phone. As e-learning has been converted to u-learning, it is expected to change to IoT-learning. This study analyzed IoT and learning system in accordance with the changes, applied the new concept of "IoT-learning", and investigated the current situation and the case to study.

Keywords: IoT · Learning-systems · U-learning · Smart-learning

1 Introduction

Nowadays, the Internet is not only familiar to us via PCs but also is used as a concept that includes the network for the various types of computer devices to communicate and share data being connected to each other at all times. The Internet of Thing products are developed like this, they have been used naturally in everyday life [18]. Turning off the light or raising the temperature in the house in the situations where no one is at home using mobile phones has already been commercialized. But there has not been a technology to make it possible for the users to receive necessary education as learners by providing the users customized data through exchanging data with all objects due to the connection of Internet [8]. So far, the learning systems such as the learning through ubiquitous technology and smart-learning have been constantly evolving. However, the researches on U-learning have not gone far off the definition of the concept stage and have remained ever in the scenario-like suggestions about the learning environment according to the ubiquitous education [8]. The existing studies have not been able to specifically mention about the changes in the learning environment and education services to be provided accordingly following the introduction

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_80

of U-learning, and have not been able to clearly suggest the benefits U-learning users such as teachers and learners of u-learning [18]. This paper provided the learner with the customized training through the data offered by a sensor network that is applied for IoT.

This paper defined the concept of IoT-learning and investigated the systems used for learning systems such as smart-learning and U-Learning, which have a similar meaning. In Sect. 3, the cases which could be used in "IoT-learning" were presented and analyzed.

2 Related Research

2.1 Learning Systems

(1) U-learning

Beyond the e-learning that is an online education simply based on internet experience, U-learning means the learning system that is possible anytime, anywhere, and to anyone in convenient manners [12]. If e-learning is an internet-based online learning, U-learning is a crystal of e-learning, which is the online learning in the U-learning environment with the availability of accessing the knowledge and information without the constraint of time and space via Any Network, Any Device [13]. With U-learning, it is possible to make use of everywhere in the world as a learning space being freed from the physical limitation of classroom by taking advantage of wireless Internet, Augmented Reality and Web reality technology including the things that exist in everyday life, the things that exist in the learning activity space, as well as the sensors, chips, and labels, and it is also possible to have personalized and customized learning according to the student interest, preference, learning styles, and learning contexts in intelligent learning environment [13].

(2) Smart-learning

Smart-learning commonly refers to the learning content and solutions which used mobile devices such as media, tablets, and e-book devices. It provides not only an internet access but also the service differentiated from the existing e-learning by taking advantage of the merits of smart devices which can be applied to a variety of techniques such as location-based services and augmented reality [6]. Sometimes smart-learning is misunderstood as the learning that takes place in various smart devices such as smartphone, tablet PC and smart TV because the term was made when the smartphone became popular, however, it is more accurate to look at the smart-learning as the learning utilizing smart devices in order to overcome various constraints which did not allow the access to learning due to technical limitations [11]. In other words, it means the people-oriented learning method where the machines support learning method of people instead of people getting adjusted to the machines and the prescribed means of learning [5].

3 IoT-Learning

3.1 IoT-Learning

(1) Definition

If the concept of IoT-Internet is that each thing with a built-in sensor makes a new value by being connected to each other, IoT-learning means to provide users with the training from something that made a new value.

It means not only to provide users with customized information by transmitting data over the network and by receiving the information to correct and analyze it, but also to provide the users with training with the information that matches the customized information. IoT-learning means to provide users (learners) with the information by collecting the information via sensors if there is education needed by the users any time anywhere. The biggest goal of this learning is to give customized training to the users with the data obtained via sensor technology of IoT. Figure 1 shows the paradigm shifts of learning from e-learning to U-learning, then to smart-learning, and expects that the learning system in the future will shift to the learning system utilizing IoT. IoT-learning is a concept created by combining the concepts of the Internet of Things, u-learning and sensor networks.

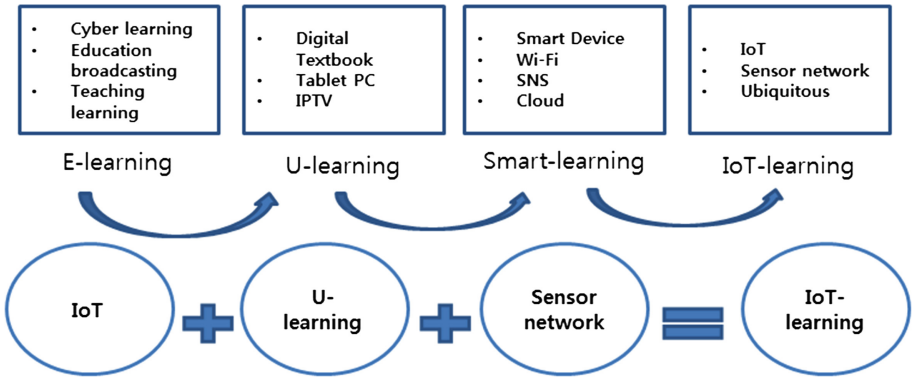


Fig. 1. IoT-leaning

(2) Differences between IoT-learning and smart-learning

The most important feature of IoT-learning is that it uses a sensor technology. As explained in the researches related to smart-learning smart-learning is limited to the things used in smart devices while IoT-learning is not limited to smart devices, also IoT-learning is different from smart-learning in the regard that IoT-learning is possible in all the devices used in both smart devices and Internet of Things.

(1) IoT-learning applied in rehabilitation

Enter the postures the patients receiving rehabilitation care must take heed into the devices such as smartphone. Give alarms to the users that it is not good to maintain their present posture any more by detecting the posture with sensors. If the alarm lasted for a few times, provide the users with a training that may help the back rehabilitation while watching the videos to be played on the display screen such as smart TV, laptop, smartphone which the users can watch right away (Table 1).

Table 1. Differences between IoT-learning and smart-learning

	Smart-learning	IoT-learning
Application of sensors	Not used	Used
Devices	Limited to smart devices	All devices

(2) IoT-learning applying Healthy Care

This is the chair using the Health Care, it provides users the stretching they need by looking for the areas of body with, for example, 'shoulder tightness' which need stretching via sensors and by analyzing the collected data anomalies by making use of the sensors attached on the chair that confirms the state of the body and can find the abnormalities of the body. The customized stretching can be offered to users by the transmission of data to the display screens in close proximity of the users such as TV or smart phone.

(3) IoT-learning applied to PT (personal training)

Time factor is something the workers who commute to work cannot overlook in considering doing a workout always in the gym with a fitness trainer. Grafting IoT-learning on PT, the users get the training on their body parts to be worked out through the videos the health trainer has already produced, perform repeat exercises, and the alarms such as continuous vibration help the user's perception in order not to have disturbance of posture by the posture sensor's continuous detection of disturbance of postures. Thus the users can continue the exercise in a more efficient way.

Table 2. Application fields of IoT-learning

Field of application	Application methods	Application display
IoT-learning for rehabilitation	(1) Enter the postures for the rehab patients to take heed. (2) Detect the postures to take heed by sensors and give alarms to the patients if the same postures last. (3) Play the stretching video on the display screen in order for the patients to stretch when the alarm lasted a few times.	Smartphone Tablet PC Laptop Smart appliances Wearable Smart Car

(*continued*)

Table 2. (*continued*)

Field of application	Application methods	Application display
IoT learning utilizing healthy care	(1) A user sit on the chair with sensors attached which can find abnormalities in the body. (2) Once the user sits on the chair, the sensor finds where stretching is needed such as 'shoulder tightness' by analyzing the user's body and inform the user. (3) Even if a user does not manually search for the video associated with stretching, offer a personalized video the user requires automatically	Smartphone Tablet PC Laptop Smart appliances Wearable Smart Car
IoT-learning of PT (personal training)	(1) Users can proceed with exercise they want to do watching videos. (2) Sensors inform the users of wrong postures while they are watching the videos. (3) It helps the users to work out with right postures.	

4 IoT-Learning Application Cases

4.1 IoT-Learning Application Cases

This study investigated the applications of IoT-learning targeting 35 patients who needed rehabilitation. People who needed rehabilitation included the patients with back pain, shoulder pain and neck disc etc. Among the long-term and short-term pain, this study investigated mainly patients with back pain. First, people with back pain have a lot of difficulties in life. Standing or sitting for a long time can give strong pressure to the back. When feeling back pain, people try to alleviate the back pain by stretching. At this time, the wrong stretches can rather worsen the back health. 20 out of 35 people undergoing rehabilitation reported that they were doing self-stretching for back pain relief. However, 15 out of 20 people were doing wrong stretching, and it rather hurt their back health. With a consideration of these situations, when the people kept maintaining the posture that might cause back pain, or worked in the same posture for a long time, the warning alarm would inform the users of that they were acting something to cause their back pain.

Vibrations via devices such as mobile phones informed the users of their wrong behaviors which might burden their backs or aggravate the pain after the reception of the data such as the postures of the patients with back pain and sitting time. For example, if $1 \sim 37$ times of vibrations through a mobile phone informed the user, the user can anywhere and anytime get the education to relieve the back pain through the display screen connected to network using IoT-learning. Rehabilitation users need the exercise to relieve the back pain by keeping the same position for a period of time. Therefore, the IoT-learning trains the users on how long they keep the same stretching. Different exercise methods are needed depending on pain and symptoms. The results of analysis based on the basic stretchings are shown in Table 2.

Table 3. Application fields of IoT-learning

Number of alarms (vibration)	Fitness level & Exercise	
1 ~ 37	Beginner	(T-L1)
38 ~ 68	Middle	(T-L1) + (T-L2)
69 ~ 80	High	(T-L1) + (T-L2) + (T-L3)
81 ~ 100	Top	(T-L1) + (T-L2) + (T-L3) + (T-L4)

Provide the users with stretching videos (SV) suitable to their back pain by increasing the strength of exercise based on the number of alarms and by offering the step-by-step exercise as in Table 3. The users can exercise step by step as in SV-L1 ~ SV-L4, and the way to do is, the more the number of alarms, the more the amount of exercise for the users to do (Table 4).

Table 4. Step-by-step example of stretching video for back pain rehabilitation

Stretching video (SV)	Name of exercise	Amount of exercise		Effect of exercise
SV-L1	Body extensor	15 ~ 30 s 2 ~ 4 times		Hip extensor stretching, such as the erector spine muscle and iliocostal muscle
	Body flexor			Rectus abdominis muscle and the internal and external oblique stretching
SV-L2	Body outside flexor	15 ~ 30 s 4 ~ 6 times		Internal and external oblique and latissimus dorsi stretching
	knee to chest			Waist and hip extensor Stretching
SV-L3	bent knee roll	15 ~ 20 s 6 ~ 8 times		Buttocks and torso and lower body muscle Stretching
	Pretzel			Internal and external oblique and hip extensor and hip piriform stretching
SV-L4	Quadrupeds lower trunk	1	10 ~ 15 s	Hip extensor stretching
		2	About 5 ~ 10 s	
	4point kneeling	15 ~ 30 s about 6 ~ 8 times		Hip extensor stretching and hip shoulder joint flexor etc. stretching

5 Conclusion

This study defined the concept of IoT-learning that provide users with customized training needed at all learning terminals connected to the networks anytime and anywhere in the perspective of ubiquitous, presented the application plans which could be utilized, and analyzed the application methods. This study defined a new concept called IoT learning grafting U-learning system on IoT technology, and analyzed based on the

cases which were applicable. As for the follow-up papers regarding IoT-learning, it is in the prospect of planning and proposing a platform regarding how to apply a learning system in reality using IoT platform by designing IoT-learning platform.

References

1. Ashton, K.: That internet of things. RFID J. **22**, 97–114 (2009)
2. Lee, J., Kim, Y.T., Lee, S., Kim, T., Choi, J.: Smart learning system using mobile augmented reality. Inf. Secur. J. **11**(6) (2011)
3. Hong, U., Eom, Z.: Case study of smart home UX design in the internet of things environment. J. Digital Des. Res. **15**(4), 3–44
4. Cho, Y.: IoT-based fusion personalized diet recommendation system framework. J. Korea Inst. Fusion Sci. **5**(4)
5. Choi, S.H.: Enabling research in e-learning contents through edutainment. Master's degree thesis at Seoul National University of Technology (2007)
6. Korea Information Society Agency: Future City Looking through Smart City. IT & Future Strategy, no. 13 (2010)
7. Choi, B.: Build a home network and U-City of Ubiquitous Era. Korea Land Hous. Corporation Land Technol. **18**(2), 87–108 (2005)
8. Son, M., Park, S., Lee, Y.: Strategy for applying spatial information services for implementing smart society. Korea Inf. Soc. Agency CLO Rep. **9**, 1–28 (2010)
9. The Gamification Summit: Day1 Conference Sessions - Make it GAMEFUL, Jane Mcgonigal (2011)
10. Essential Facts about the computer and video game industry, ESA (Entertainment Software Association) (2011)
11. Reality Is Broken: Why Games Make Us Better and How They Can Change the World, Jane McGonigal (2011)
12. Lee, H., Kim, M., Bang, H.: Internet of things technology trends and future direction. Inf. Process. Soc. **21**(2) (2014)
13. Connecting Lab: Huge Connection beyond Cloud and Big Data, IoT. Window to the Future (2014)
14. New Media Consortium: The NMC Horizon Report: 2011 K-12 Edition. emqusa.com (2011)
15. Sasaki, S., Watagoshi, K., Takano, K., Hirashima, K., Kiyoki, Y.: Impression-oriented music courseware and its application in elementary schools. Interact. Technol. Smart Educ. **7**(2), 85–100 (2010)
16. Cho, S.: Design and implementation for smart learning system for listenig to music. Music Educ. Eng. **1**, 231–249 (2014)
17. Oh, J., Kim, H.: A study on application of virtualization for U-learning system optimization. Korea Inf. Sci. **6**, 236–240 (2010)
18. Yun, D., Ju, J.: A study on foundation technologies for implementation of ubiquitous learning system. In: Proceedings of Business Engineering, vol. 11, pp. 668–675 (2005)
19. Park, C., Hong, Y.: A study on the implementation of U-learning. Korean Electron. Eng. **6**, 790–791 (2012)

Author Biographies

Seul-Ah Shin obtained her Bachelor's degree in the area of Computer Science at Korea National Open University, is currently in the master's program in the Department of Computer Science at Soongsil University. The areas of main interest are Master of U-learning system, Internet of Things, software quality management etc.

Ji-Soo Choi obtained his Bachelor's degree in School of Computing at Soongsil University, is currently in the master's program in the Graduate School of Information Science at Soongsil University, and The Korea Society of Information Technology Policy & Management (KITPM) researcher. The areas of main interest are Master of Software Engineering, Internet of Things, software quality management etc.

Young-Jong Kim obtained his Master's degree in Engineering at Soongsil University, and Ph.D in IT Policy at Soongsil University. He is currently serving as a professor in the Department of Software at Soongsil University, and served as an examiner in open software contest, embedded software contest and Korea, China, Japan open software forum. The area of main interest includes network computing and Scalable Application Architecture etc.

Nam-Yong Lee obtained his Bachelor's degree in Computer Science at Soongsil University, a Master's degree in Business Administration in the area of MIS, at the Graduate School of Management Information (MIS), Korea University, and Ph.D. in Business Administration in the area MIS at MSU (Mississippi State University), served as KIDA senior researcher, KIDA/ADD research director, and the chairman of The Information & Communication Professional Engineers Association of Korea, and is currently serving as a professor in the Department of Software at Soongsil University. The area of main interests include software testing, QA, MIS, IT policy management.

Jin-Ho Park obtained his Bachelor's degree in Software Engineering at Soongsil University, and Master's and Ph.D in Software Engineering at Soongsil University. He is currently serving as a professor in the Department of Software at Soongsil University, and the areas of main interests include SW Safety/Quality/Testing, SW Fusion/soft power, Internet of Things, military ISR, IT Service, IT technology commercialization/ Start-Up.

A Theoretical Study of Hardware Architecture for Network Security Server

Joong-Yeon Lee[1], In-Taek Oh[2], Nam-Yong Lee[3], and Jin-Ho Park[3(✉)]

[1] KTNF, Seoul, Korea
realtime@ktnf.co.kr
[2] KT, Seongnam, Korea
itoh@kt.com
[3] School of Software, Soongsil University, Seoul, Korea
{nylee, j.park}@ssu.ac.kr

Abstract. Development of IT led to changes in the Infra architecture and the area of information security has been especially emphasized. With the rapidly growing Data, the area of Information Security has established itself as an essential, and the Network Security Field is expected to be developed and elaborated also for the private and public peace and the establishment of social order. The purpose of this study is to consider the architecture of high-performance hardware architecture that can process large amounts of Data, and also to study the architecture of the hardware with flexible architecture in accordance with the amount of network traffic.

Keywords: Hardware architecture · Information security · Network · Security server

1 Introduction

Recently, IT industry, which has been developed rapidly, has been producing large amount of DATA, and the demands for real-time processing are on the rise. In particular, the ISSUE of DATA security of network terminals are on the rise, while the real-time processing is a critical problem and is exceeding the limitation of DATA which can be processed on a single server.

In the last 20 years, owing to the advances in semiconductor process technology, the computing power increased exponentially, but it has been presumed to be impossible to overcome the limitations of utilizing a single CPU compared to the latest status of DATA flooding. There can be a variety of approaches to ensure the computing power that can process large amounts of high-speed processing in real time by incorporating the hardware and software technology. This paper proposed a method that could expand the computing power of single CPU with flexibility.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_81

2 Cloud Computing and Learning System

This study proposed a hardware architecture capable of high-capacity high-speed processing by analyzing the architecture of existing single computing node and by expanding the small single computing power. Among the numerous existing CPU manufacturers, this paper analyzed the architecture of the computing module utilizing the x86 CPU family of Intel Corporation with the highest market dominance, performed the analysis of the minimum requirements for each computing module, and reviewed the available specifications for cross-DATA INTERACTIONs among Computing Modules as well as the existing technologies which have been SWITCHING the high-capacity high-speed DATA [1].

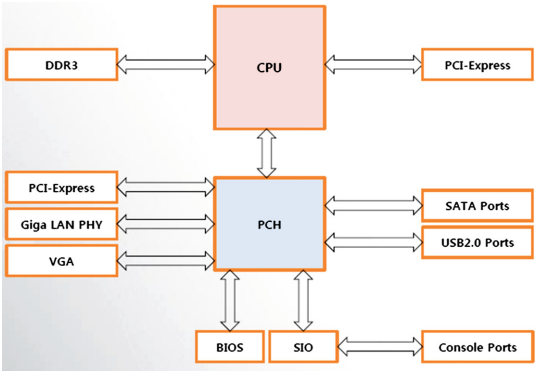


Fig. 1. A study for a single computing module architecture

2.1 Structural Analysis of Computing Module

Figure 1 shows a block diagram of computing module of x86 CPU of Intel Corporation used universally. A single computing module on a large scale consists of a CPU and PCH (Platform Control Hub), where PCH plays a role of HUB of various IO, disk, and system interface. It is important to consider the operating frequency of CPU, size of internal Cache memory and bandwidth of external memory in order to raise the performance of a single computing module as much as possible. Recent technology trend is mainly the technology of many core CPU that resolves the side effect of heat generation by applying many CPU cores in a single package and by raising the operating frequency via the sharing of Last layer Cache by each CPU core. In addition, one of the new changes is the change from the parallel data bus architecture to the serial bus architecture. In the case of the parallel bus architecture, it has the architecture sensitive to the signal delay or the skew of signal, and in order to transmit a lot of data from the same clock, the number of signals increases, thus occupies a lot of space and lead to the greater the power consumption. In order to solve many of these problems, data bus was serialized, its typical standardized interface is PCI Express, then Gen1, Gen2, Gen3 standard have been released, and in the case of Gen3, 8 GHz Data

communications is ensured over a single Lane. PCIe, which ensures low latency and high Bandwidth has been adopted by most of the CPU's including intel x86 CPU, as well as has been used as a standard IO devices such as Ethernet Controller, Disk controller, and Video Controller [2].

2.2 Research on Data Switching

Each single computing module is in need of a route for transmitting the results of the calculation of data. In analyzing the process of data transfer, a transaction should be made in the format which can be recognized to each other, and this is called a protocol. There can be many ways to perform data communication between the computing modules, but the reliable methods to transmit a large amount of Data at a high speed are the Ethernet communication and the method utilizing PCIe NTB (Non Transparent Bridge). The advantages of data transmission utilizing Ethernet Switching Protocol is a stable protocol and the fact that long-time proven software stack are secured. PCIe switching technology utilizing NTB also supports low packet latency and high bandwidth. The switch fabric block of Figs. 2 and 3 can be an Ethernet switch or the PCIe switch [3, 4].

2.3 Considerations of Power Consumption and Heat Dissipation

The biggest byproduct of high frequency clocking is the generation of heat and power consumption. All the electronic components require power and, in the case of high Clocking, more power consumption and heat are generated. With the development of semiconductor process technology, the power consumption was reduced, but it is not a fundamental solution, and the consideration for power and heat generation is required at the system level. Just as there is a limitation on the power produced by power stations, there is a limit to the power supplied to the building in the cases of the Server farms and data centers to process large amounts of data. In addition, the same applies to the power provided to a single Rack [5].

3 Hardware Architecture for High-Performance Network Security Server

In order to implement the Extreme High-End computer system, it is possible to implement a High performance Network Security Architecture when mutual harmony for the same purpose is achieved than the separate point of view of hardware and software. Depending on the nature of Network Security, it is either necessary to have realtime In-Line packet processing architecture design in terms of an application, or, it is possible to have a Look Aside architecture that transmits first and then performs the packet test separately. However, in the case of the application with critical Mission, the real time in-line architecture to detect the malwares or attacks by inspecting every packet real time is suitable. In addition, it is also possible to design a sequential

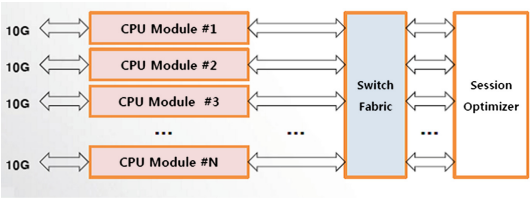


Fig. 2. Data path interconnection of multiple computing modules utilizing the high-speed serial standard Fabric

architecture that divides the same packet into the pre-treatment and post-treatment by looping via the division of roles among cpu modules in order to overcome the limitations of performance. It is important to have a flexible hardware architecture design that can be scaled out to meet these diverse requirements.

In addition, for high performance, it must have an interface specification that is uniform among multiple CPU Modules and can be connected at low latency, which can overcome the limitations of a single CPU, and each CPU Module may be either the same type or the different type. Through the Unified Standard Interface, it is possible to secure the Scalability, Portability, and Flexibility. Currently, the available System to System interfaces are SRIO, PCIe, Ethernet, Infiniband etc., among them, the Ethernet and PCIe are mainly used.

Although there may be a variety of researches to give a way to overcome the limitations of the processing capacity of a single computing module, Figs. 2 and 3 show the examples of data path interconnection of multiple computing modules utilizing the high-speed Switching Fabric.

In the case of Fig. 2, multiple CPU Modules are connected through standard Interfaces with Switch Fabrics, and each CPU Module is connected to the external Network via a separate Ethernet. At this time, the incoming Ethernet packet will be processed on the CPU Module based on security policies, and also the resources needed to be shared by each CPU Module can be implemented to be managed at the Session Optimizer through the Switch Fabric.

The switch fabric can be either a PCIe switch or an Ethernet Switch, and the switching chip should be selected in consideration of the sum of the performance of each CPU Module. In this study, the Ethernet Interface of CPU Module is schematized to be 10 Gbps, but the 40 Gbps performance of a single CPU Module can be expected given the recent performance of CPU.

In the case of Fig. 3, it may look similar to Fig. 2, but it can be seen as the opposite case in Handling Method of Packets. In the case of Fig. 2, it shows the architecture where the CPU Module receives a Packet first at the Edge terminal to perform the pre-processing, but in Fig. 3, it is limited to the architecture using the Ethernet switch, and the packet entered the switch fabric transmits the result after the packet Pre-processing at a switch according to security policies. A Session Optimizer can be used as a control plane of policy switch fabric that is treated with session management and CPU Module.

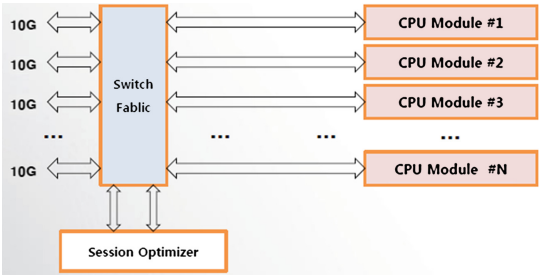


Fig. 3. Data path interconnection of multiple computing modules utilizing the high-speed Ethernet Switching Fabric

4 Conclusion

This study is about the hardware architecture of network security server with its available reliable high-speed packet processing against the threats to information security which have been becoming increasingly intelligent and elaborate abreast of the development of information technology. It is presumed, through this study, that a clue to search for an answer among many possibilities can be given, and that new studies on architecture can continue and be developed.

References

1. Vinogradov, V.I.: Advanced high-performance computer system architectures. Nucl. Instrum. Methods Phys. Res. A **571**, 429–432 (2007)
2. Migliardi, M.: Performance evaluation of the SM-IMP architecture: a parallel, heterogeneous, image processing oriented architecture. In: RAS, Scientific Instrumentation, Band 5, N3-4, S-Petersburg (1995)
3. Bianco, A., Giaccone, P., Ricca, M.: Scheduling traffic for maximum switch lifetime in optical data center fabrics. Department of Electronics and Telecommunications, Politecnico di Torino, Italy. E-mail: andrea.bianco@polito.it, paolo.giaccone@polito.it, marco.ricca@polito.it
4. PCI Express® Base Specification, Revision 3.0, 10 November 2010
5. Frankiewicz. M.: Overheat protection circuit for high frequency processors. Bulletin of the Polish Academy of Sciences. Technical sciences

Author Biographies

Joong-Yeon Lee obtained his Master's degree in Electrical and Electronic Engineering at Yonsei University Graduate School of Engineering, has served as the CEO of KTNF Inc. since 2001, and has been developing and manufacturing computer servers. The areas of main interests include network security server, Open Compute Project, Software Defined Network, ALL FLASH STORAGE development, New Architecture Design etc.

In-Taek Oh obtained his Master's degree in Economics in the area of information technology economics at Graduate School of Economics, Sogang University, is currently in the PhD program of Soongsil IT Policy Business Administration, served as a consultant and PM (Project Manager) for a number of large-scale projects, and is currently (May 2016) overseeing kt's ERP, OSS, BIDW in kt IT Planning Office. The areas of main interests are MIS, ERP, BigData, object-oriented software engineering/UML, and Cost effective project methodology.

Nam-Yong Lee received his bachelor's degree of Computer Science in Soonsil University, Seoul and master's degree of MIS in Graduated Korea University, Seoul. doctor's degree of MIS in Mississippi State University, USA. Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on System Engineering, Software Engineering, E-Commerce System, MIS, etc.

Jin-Ho Park obtained his Bachelor's degree in Software Engineering at Soongsil University, and Master's and PhD in Software Engineering at Soongsil University. He is currently serving as a professor in the Department of Software at Soongsil University, and the areas of main interests include SW Safety/Quality/Testing, SW Fusion/soft power, IoT, military ISR, IT Service, IT technology commercialization/Start-Up and more.

An Empirical Study of the Relationship Between DISC Behavioral Style of Application Programmer and Quality of Software Development

In-Taek Oh[1], Joong-Yeon Lee[2], Jae-Yoon Cheon[3], Nam-yong Lee[4], and Jin-Ho Park[4(✉)]

[1] KT, Seongnam, Korea
itoh@kt.co
[2] KTNF, Seoul, Korea
realtime@ktnf.co.kr
[3] LG NSYS, Seoul, Korea
jycheon@lgnsys.com
[4] School of Software, Soongsil University, Seoul, Korea
{nylee, j.park}@ssu.ac.kr

Abstract. A number of application programmers' factors influence the quality of software development. The purpose of this study was to examine the relationship between the DISC behavioral styles of application programmers and the quality of software development for identifying how the factors affect the quality of software development. We conducted a field experiment with 34 application programmers working for a certain company. In conclusion, it may be said that the quality of software development is affected by the working periods of application programmers rather than their DISC behavioral styles.

Keywords: DISC behavioral style · Application development · Software quality

1 Introduction

1.1 Necessity of the Study

Necessity of What is the most important of all in performing a project is a developer's compliance with coding standards. Each company pays much attention to this, for in case of further advancing or maintaining a project, additional huge maintenance cost may be incurred depending on the compliance with standards.

Even when coding standards for a relevant project have been established, however, the standard compliance rate varies with developers involved, and although a lot of management staff members are allocated for the inspection of the compliance rate, there are many cases where actually developed source code fails to comply with coding standards.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_82

Developers' will to comply with the standards is most important of all, and the compliance rate tends to change considerably according to ABAP developers' characteristic behaviors.

Behavior is an individual's external traits, and all individuals show different behavioral styles as they grew in different environments. Dr. William M. Marston, professor at the US Columbia University, modeled four human behaviors of DISC on such behavioral styles.

He named the four human behavioral patterns Dominance (D), Influence (I), Steadiness (S), and Conscientiousness (C). His DISC theory has been translated into 17 languages in 55 countries across the world, and is used to maximize business performance at the scene of management or to motivate relevant members [GLS worldwide 2011].

Each developer's DISC behavioral style is easy to measure, and the DISC assessment costs little money. Given this, this study tries to present basic data for creating cost-effective methods for allocating developers and managing the allocated developers, by applying the DISC assessment to project management, with focusing on compliance with project coding standards the study.

1.2 Purpose of the Study

This study intends to investigate whether application programmers' DISC behavioral styles and their five external factors, that is, working period at the same site, age, overall IT experience, and academic background, are correlated with the quality of software development, respectively. The purpose may be further specified as follows:

First, programmers' compliance with standards according to their DIC behavioral styles and its correlation with program execution error will be analyzed.

Second, programmers' apparent traits and compliance with standards, and their correlations with program execution error will be analyzed.

1.3 Definition of Terms

1.3.1 DISC Behavioral Styles

The theory of DISC behavioral styles was developed in 1928 by William M. Marston, professor in psychology at Columbia University, who classified human behavioral types into Dominance (D), Influence (I), Steadiness (S), and Conscientiousness (C). This study used the behavioral types classified by the Korean-version of 'DISC Inventory' translated by the Korea Educational Consulting Institute (2016), the exclusive distributor of DISC PPS (personal profile system) in Korea.

1.3.2 Application (Program)

It does not imply that database is built or that components are created, but it is developed so that a program having the user's desired functions may be provided between the GUI layer and the application layer according to business process.

1.3.3 Coding Standards for Development

'Naming Rule' and 'Whether comments written or not' were defined as the coding standards. The 'Naming Rule' comprises program name, function name, table name, and data element name. 'Whether comments written or not' is to describe in the main logic the reasons and grounds for programming. An ordinary project defines the standards suitable for a relevant site, and programmers are offered guidance before being allocated to the project so that the programmers may be fully acquainted with the standards and then develop a program.

1.3.4 Rate of Non-compliance with Standards

It means the number of objects that fail to comply with the standards among objects generated by a programmer.

(1) Rate of non-compliance with programming standards (%) = object(s) of non-compliance with the naming rule / all objects generated by a programmer * 100.
(2) Rate of comments not written = object(s) without comment / all objects generated by a programmer.

2 Related Studies

As for DISC-applied studies in various areas, most of them deal with the relationships between DISC behavioral styles and leadership, business performance, job satisfaction, job stress, or organizational effectiveness [Kang 2007; Kim et al. 2006; Kim 2012; Kwak, 2012]. While there is a study in the area of medicine that researched about the correlation between the DISK styles and medication error [DISC Behavior Pattern and Medication Errors by Nurses 2013], in the area of software, there has been no research about the relationship between DISC and the quality of application and software.

3 Research Methods

3.1 Research Design

This study is field research using a field experiment. To identify the relationship between the DISC behavioral styles of application programmers allocated to a specific company and their non-compliance with program coding standards, this study examined all source codes of programs actually created by programmers in order to find out non-compliance with standards, by accessing the operation system together with two programming experts. In addition, to find out error in operating system, the dump from January 2014 to April 2016 was analyzed; whether having been written by programmers, who were then allocated, was detected; and the correlations were investigated.

3.2 Subjects of Research

All 34 application programmers working for a specific company were surveyed. Programs and objects generated by the programmers, and errors resulting from them during a period between January 2014 and April 30, 2016 were investigated.

3.3 Methods for Data Collection

The purpose of the study was explained to all the 34 application programmers of the specific company. First, to diagnose the programmers' behavioral styles, a written survey of the programmers was carried out, using the 'DISC personal profile system.' Then, on the basis of the programmers' resumes obtained from the personnel department of the specific company, their academic background, overall IT experience, and working periods were identified. Programs, which were developed by the relevant programmers for the actual operating system, and programs in which error occurred were extracted. In addition, programmers were sorted out whose program codes didn't comply with the standards of 'Development Standards Handbook' and 'AP Quality Control System,' which are currently applied to development business affairs.

3.4 Research Tool

(1) Number of erroneous programs

It means the number of programs whose error was not detected by their programmers during their development and testing, but was found during the actual execution of the programs. Given the possibility that the same erroneous program may be executed several time and the same error case be counted repeatedly, the error was counted once a day for the same program.

(2) Development Standards Handbook

It is a guidebook that defines the Naming Rule for every object, and is offered to programmers assigned to application development. Each programmer should become well-acquainted with the guidebook before creating a program. The guidebook also serves as a criteria tool for an inspector to check whether naming standards were correctly complied with. The Naming Rule was defined for a total of 28 object types. 'Development class', 'Programs', 'Module Pools', 'Function Group', 'Function Modules', 'Transaction Codes', Tables', 'Table Index', 'Table Field', 'Data element', 'Message Class', 'Structure', 'View', 'Domain', 'Lock', 'Object', 'Range', 'Internal table', 'Variable', 'Constant', 'Parameter Field', 'Select-option', 'Area Menu', and 'Module Code' define the meaning of each digit for objects that must be created or may be created when a programmer writes a program.

(3) Whether program comments written or not

Every programmer should add explanations to the source codes of their own programs as comments, as shown in 'Fig. 1. An illustration of comments.' Out of an application created by each programmer, 10 source codes are randomly chosen and checked about whether comments were written or not, and in case of any source code

```
"주민번호가 존재하는 경우 생년월일, 성별로 분리
  IF gs_item-stcd1 IS NOT INITIAL.

    IF gs_item-ktokd EQ 3300.  "외국인
      gs_item-stcd1_f = gs_item-stcd1.

    ELSE. "외국인 이외에는 주민번호로 분류
      gs_item-stcd1_b = gs_item-stcd1+6(1). "성별구분

      CASE gs_item-stcd1_b.    "성별 구분자
        WHEN c_1 OR c_2.  "1900년대
          CONCATENATE '19' gs_item-stcd1(6) INTO gs_item-stcd1_a.
        WHEN c_3 OR c_4.  "2000년대
          CONCATENATE '20' gs_item-stcd1(6) INTO gs_item-stcd1_a.
        WHEN c_9 OR c_0.  "1800년대
          CONCATENATE '18' gs_item-stcd1(6) INTO gs_item-stcd1_a.
        WHEN OTHERS.
      ENDCASE.

    ENDIF.

  ENDIF.

  APPEND gs_item TO pt_item. CLEAR : gs_item, gs_kna1.
```

Fig. 1. An illustration of writing comments

with no comment, the program is judged to have not complied with the standards. Three development experts analyzed the contents of comments for 10 source codes, and classified as No Program Comment Written in case of incorrect description.

3.5 Methods for Data Analysis

The collected data were analyzed, using the Microsoft Excel program. From this, the number of programs created, the mean of IT experience years, and the mean rate of error were calculated according to programmers' behavioral styles. The correlations between the behavioral styles and the error rate were analyzed, and the environmental characteristics of the programmers were analyzed.

4 Results of the Study

4.1 Compliance with the Standards According to the Programmers' Behavioral Styles, and Its Correlations with Program Execution Error

See Table 1.

Table 1. Program error rate by DISC behavioral style

Behavioral style (n)	Total number of programs generated (copy)	Number of programs with error (copy)	Error rate (%)
Type D (3)	179	3	1.7 %
Type I (11))	4,265	170	4.0 %
Type S (9)	3,165	175	5.5 %
Type C (11)	3,239	53	1.6 %

In addition, results of investigating whether objects of each program complied with the Naming Rule are as shown in Table 2.

Table 2. Compliance with Naming Rule according to DISC behavioral styles

Behavioral style	Status of Naming Rule compliance (rate of compliance)
Type D	All the three complied (100 %).
Type I	Nine of 11 complied (82 %).
Type S	Four of nine complied (44 %).
Type C	Nine of 11 complied (82 %).

In addition, results of investigating whether objects of each program complied with the Naming Rule are as shown in Table 3.

Table 3. Status of comments writing according to DISC behavioral styles

Behavioral style	Status of comment writing (rate of writing)
Type D	All the three wrote (100 %).
Type I	Three of 11 wrote (27 %).
Type S	Two of nine wrote (22 %).
Type C	Four of 11 wrote (36 %).

4.2 Analyzing the Apparent Traits and Standards Compliance of Programmers, and Their Correlations with Program Execution Error

The apparent traits of programmers comprised overall IT experience, working period at the same site, and academic background, among which the academic background was excluded because all except one graduated from college. First, the program error rate against the overall IT experience is presented in Table 4. As shown in Table 4, the longer IT experience does not necessarily imply positive (+) or negative (−) relations, and thus it may not be said that there is a correlation between the overall IT experience and the program error rate.

In addition, the relationships between the working period at the same site and the error rate are presented in Table 5. As shown in Table 5, the program error rate tends to be high at 1 year and 6 months after a programmer was allocated, thereafter abruptly

Table 4. The program error rate (%) against overall IT experience and the total number of program copies

IT experience	Below 5 years	Below 10 years	Below 15 years	Below 20 years	Below 25 years
Error rate	3 %	3 %	13 %	2 %	1 %
Number of programs	1,797	6,380	868	1,628	175

decreasing. In particular, it is estimated that the reason for the high error rate at 1.5 years after allocation is because many programs of a high level of difficulty were created compared to below 0.5 years or below 1 year.

Table 5. The working period at the same site and the program error rate (unit: year)

Work years	Below 0.5	Below 1	Below 1.5	Below 2	Below 3	3 or over
14 %	14 %	10 %	43 %	1 %	3 %	2 %

4.3 Analyzing the Combination of Programmer's DIS Behavioral Styles and Apparent Traits, and Its Correlations with Program Execution Error

The DISC behavioral styles of programmers, the working period at the same site, and the program error rate are presented in Table 6. As shown in Table 6, it was found from its analysis that the program error rate of programmers is correlated only with the working period at the same site, rather than the DISC behavioral styles of programmers. It is judged that the error rates for Type D and Type C are small because their working period is 2 years and above, compared with Type I and Type S.

Table 6. The program error rates against the working period at the same site and the DISC behavioral styles

Division	D	I	S	C
Below 0.5 years	0 %	10 %	23 %	N/A
Below 1 year	N/A	N/A	16 %	N/A
Below 1.5 years	N/A	75 %	43 %	N/A
Below 2 years	10 %	1 %	3 %	1 %
Below 3 years	N/A	3 %	N/A	1 %
Below 4 years	N/A	3 %	2 %	2 %

5 Conclusion

This study classified programmers' traits, using the tool of DISC behavioral styles, and considered whether there is any association between the styles and program error rate. The findings of the study indicated the low program error rates of 1.7 % and 1.6 % in

the case of Type D (Dominance) and Type C (Conscientiousness), respectively, and relatively high rates of 5.5 % and 4 % in the case of Type S (Steadiness) and Type I (Influence), respectively. The investigation into the relationships between the error rate and the working period showed that errors were concentrated in programmers working below 1.5 years after their allocation. Further, it was found that in the case of Type D (Dominance) and Type C (Conscientiousness), all the programmers falling under them worked two years or longer. In addition, in the case of programmers working two years or longer since allocation falling under Type S (Steadiness) and Type I (Influence), their program error rates were measured to be similar to those for Type D (Dominance) and Type C (Conscientiousness).

Therefore, it is judged that almost no program error will occur if newly allocated programmers are thoroughly educated on 'Development Standards Guidelines' and 'Development Process' system designed to intensively inspect and check programs produced by them.

In the future study, we will investigate the relationships between the DISC behavioral styles and the error rate by expanding the population and the target companies. Besides, we will investigate programmers' traits connected with program error, and prepare standards that enable companies to cost-effectively select and source good programmers.

References

GLS Worldwide: Behavioral style insight report for jane sample, 17 March 2011. http://www.glsworld.com/pdf/assessment-sample-reports/development-behavioral-style-insight.pdf. Accessed 26 Sep 2012

Hollnage, E.: Human Reliability Analysis: Context and Control. Academic Press, London (1993)

Kang, J.H.: Study on the DISC pattern of constituent and influence of revolutionary leadership that affects to job stress. Unpublished Master's thesis, Kyung Hee University, Seoul, Korea (2007)

Kwak, Y.J.: The effect of supervisor-subordinate DISC behavioral tendencies on subordinates's organizational commitment and job satisfaction. Unpublished Master's thesis, Sookmyung University, Seoul, Korea (2012)

Korea Educational Consulting Institute, 15 May 2016. http://www.kdisc.co.kr/page/sub2_1

Author Biographies

In-Taek Oh obtained his master's degree in IT economics from Graduate School of Economics, Sogang University, and is now a doctoral candidate at Department of IT Policy and Management, Soongsil University. He participated in a number of large projects as consultant and PM (project manager). He is now responsible for ERP, OSS, BIDW, and BigData at kt's IT Planning Office. His main areas of interest includes MIS, ERP, BigData, object-oriented/UML in software engineering, and cost-effective project methodology.

Joong-Yeon Lee obtained his master's degree in electrical and electronic engineering from Yonsei Graduate School of Engineering in 2011. He has served as Representative Director of KTNF, a computer server developer and manufacturer, since 2001. His main areas of interest include network security server, all flash storage development, and new architecture design.

Jae-Yoon Cheon obtained his master's degree in computer engineering from Yonsei Graduate School of Engineering, and is now a doctoral candidate at Department of IT Policy and Management, Soongsil University. He is currently on active duty as expert member at Infrastructure Division, LG N-Sys, and is participating in a number of large projects as consultant and PM (project manager). His main areas of interest include open-source software and SDN (software defined networking).

Lee Nam-Yong obtained his bachelor's degree in computer science from Soongsil University, his master's degree in MIS from Korea University Business School, and his doctoral degree in MIS from Mississippi State University. He is now a professor at School of Software, Soongsil University. His main areas of interest include system engineering, software engineering, and electronic commerce system/MIS.

Jin-Ho Park obtained his bachelor's degree in software engineering from Soongsil University, and his master's and doctoral degrees in software engineering from Graduate School of Soongsil University. He is now a professor at School of Software, Soongsil University. His main areas of interest includes software safety/quality/testing, software convergence/ soft power, IoT, defense ISR, IT service, and IT technology commercialization/start-up.

Advances in Computer Science and Ubiquitous Computing

The Development of COB Type LED Lighting System for High Temperature Machine Vision

Park Sanggug(✉)

Division of Steel-IT Engineering, Uiduk University,
Kyungju, Kyungpook, Korea
skpark@uu.ac.kr

Abstract. In the machine vision system of steel production line, the obtained image qualities are greatly affected by brightness and wavelength band of illumination. The surface temperature of a continuous casting material is very high, 600 ~ 1000 °C. But the operation temperature of COB type LED is maximum 90 °C. Therefore, any special cooling systems are needed to use in these high temperature environment. In this paper, we have developed COB type LED lighting system for high temperature machine vision equipment. Our LED lighting system consist of air filter module, heat sink module, reflector and reflecting plate, heat blocking glass, air injector and circulation module. We suggest air curtain system to heat block in front of heat-blocking glass. Through the test, we have confirmed that our air curtain system will drop the internal temperature to below 80 °C in the circumstance of high temperature 600 ~ 1000 °C.

Keywords: COB type LED · Machine vision · LED lighting system · Air curtain system

1 Introduction

The brightness of indoor lighting should be sufficient and uniform to protect eyesight and reduce eye fatigue. In the machine vision system, the light is one of the most important facts. The reliability, accuracy and response time of the vision system can be improved by selecting the appropriate lighting structures. In the machine vision system, the image acquisition processing in inappropriate lighting environments may be accompanied by a fatal error. The compensation algorithm to the inappropriate image causes a large load on the system. The computer vision or machine vision area is essential to the measuring, defect detection and process monitoring in manufacturing system. The important factors in the visual recognition of the human are color, perspective, shadow, parallax and personal experience. But, the vision system does not have the empirical basis on which to make decision, and should specify everything [1–3].

In the steel production line, surface defection does effect seriously to the next heat treatment processing and quality of final steel manufactures. Therefore, defect detection is very important. In the continuous casting process of steel production line, the machine vision system has been applied for defect detection. The obtaining a high quality images by vision camera is very important for accurate defect detection in the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_83

vision system. These obtained image qualities are greatly affected by brightness and intensity of the illumination, and wavelength band of illumination in the lighting system. Also, power consumption and life time of lighting system are very important factors. The COB (Chips On Board) type LED lighting satisfies these factors, that is high intensity illumination, uniform illumination, wide band wavelength, low power consumption and long life time. The surface temperature of a continuous casting material in the steel production line is very high, 600 ~ 1000 °C. But the operation temperature of COB type LED is maximum 90 °C. Therefore, any special cooling systems are needed to use in these high temperature environment [4, 5].

In this paper, we have developed COB type LED lighting system for high temperature machine vision equipment. Our LED lighting system consist of air filter module, heat sink module, reflector and reflecting plate, heat blocking glass, air injector and circulation module. We suggest air curtain system to heat block in front of heat-blocking glass. Through the test, we have confirmed that our air curtain system will drop the internal temperature to below 80 °C in the circumstance of high temperature 600 ~ 1000 °C (LED contact temperature is about 500 °C).

2 Theoretical Backgrounds

2.1 Machine Vision System

Machine vision (MV) is the technology and methods used to provide imaging-based automatic inspection and analysis for such applications as automatic inspection, process control, and robot guidance in industry. A machine-vision system employs one or more video cameras, analog-to-digital conversion (ADC), and digital signal processing (DSP). The resulting data goes to a computer or robot controller. Machine vision is similar in complexity to voice recognition [6, 7].

Two important specifications in any vision system are the sensitivity and the resolution. Sensitivity is the ability of a machine to see in dim light, or to detect weak impulses at invisible wavelengths. Resolution is the extent to which a machine can differentiate between objects. In general, the better the resolution, the more confined the field of vision. Sensitivity and resolution are interdependent. All other factors held constant, increasing the sensitivity reduces the resolution, and improving the resolution reduces the sensitivity. Human eyes are sensitive to electromagnetic wavelength ranging from 390 to 770 nm. Video cameras can be sensitive to a range of wavelengths much wider than this. Some machine-vision systems function at infrared (IR), ultraviolet (UV), or X-ray wavelengths. Figure 1 shows general configuration of the vision system for inspection. As show in Fig. 1, uniform light is very important for good inspection. Figure 2 shows configuration of the existing machine vision system for surface defect inspection of continuous casting slab in the steel production line. The halogen lamp, metal halide lamp and plasma lamp are used for illumination of machine vision. The disadvantages of these illuminations are short life time and easy deterioration in high temperature environment. In recent, the LED light is used to address these problems.

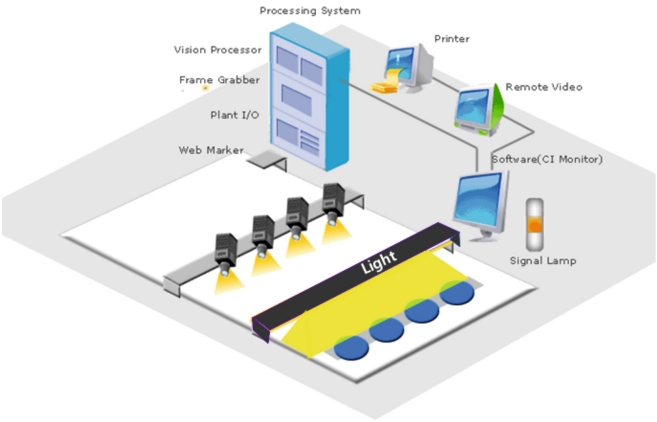


Fig. 1. General configuration of the vision system for inspection

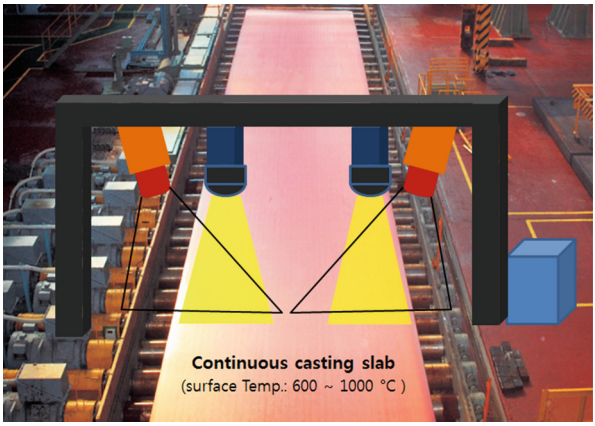


Fig. 2. Configuration of the existing machine vision system for defect inspection in the steel production line

2.2 The COB Type LED Light

COB is a new technology of LED packaging for LED light engine. Multi LED chips are packaged together as one lighting module. When it light up, it looks like a lighting pane (Fig. 3).

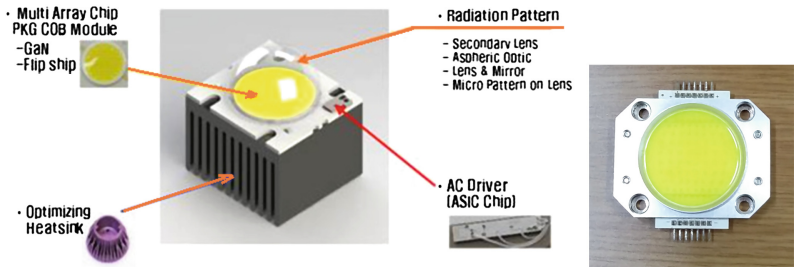


Fig. 3. Configuration of the existing COB type LED device

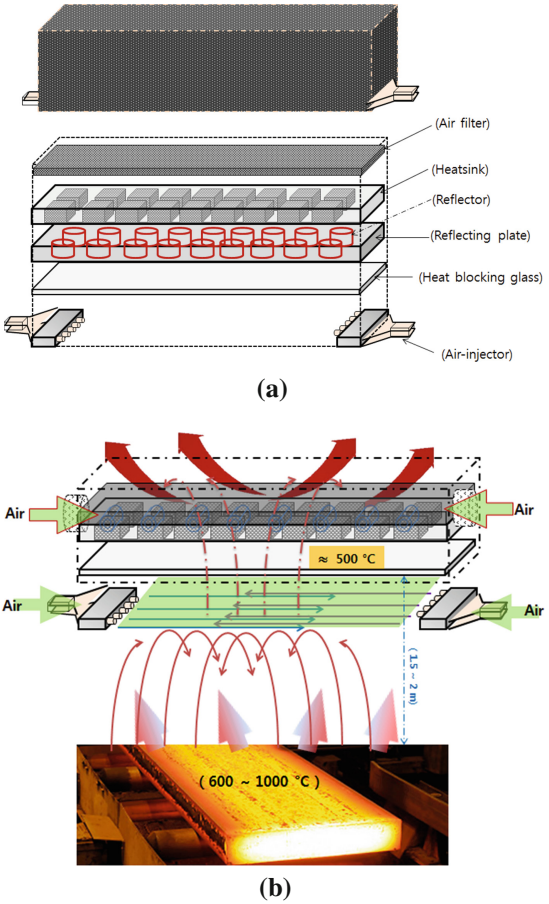


Fig. 4. Configuration of COB type LED lighting system ((a) overall system, (b) air curtain system and heat release mechanism)

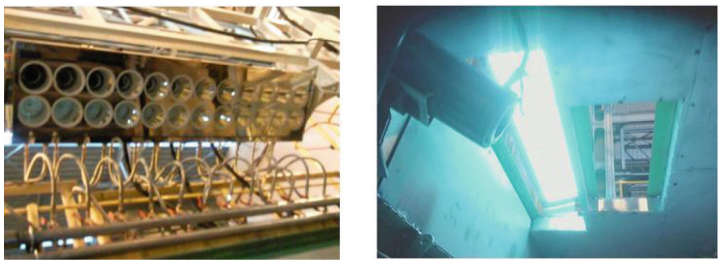
3 COB Type LED Lighting System for High Temperature Machine Vision

We have designed COB type LED lighting system for high temperature machine vision equipment. As shown Fig. 4(a), our LED lighting system consist of air filter module, heat sink module, reflector and reflecting plate, heat blocking glass, air injector and circulation module. We suggest air curtain system to heat block in front of heat-blocking glass. As shown Fig. 4(b), cold air which is injected from "air-injector" make air curtain in front of heat-blocking glass. Air curtain is located about 5 m distance from the surface of high temperature steel material (600 ~ 1000 °C). In that location, LED light contact temperature is about 500 °C.

4 Results and Discussion

Figure 5(a) shows prototype of developed LED lighting system, and Fig. 5(b) shows photo of field installed system. We have tested our system for one month in the continuous casting slab line of steelmaking process in steel & iron company, Pohang, Korea. Through the test, we have confirmed that our air curtain system will drop the internal temperature to below 80 °C in the circumstance of high temperature (LED contact temperature is about 500 °C). In the future, we will continue reliability test and life-time test of lighting system.

(a)



(b)

Fig. 5. Developed LED lighting system ((a) prototype system, (b) field installation system)

Acknowledgments. "This research was supported by academic-industrial cooperation technology development project of Small and Medium Business Administration (*No: C0397605*)".

References

1. Park, Y.-J.: Using high brightness LED light source controller for machine vision. J. Digital Convergence **12**(4), 311–318 (2014)
2. Jeong, S.-C.: Development of simulator for analyzing and designing a lighting system of machine vision. Master's thesis, Dep. of Information Engineering, InHa University (2009)
3. Park, S.-H.: A Study on LED Strobo Driver for Machine Vision System with current control function. Master's thesis, Dep. of Electronics Engineering, ChungJu University (2011)
4. Kim, J.-H., Jeong, M.-S., Kim, S.-W.: Aspherical lens design of LED surface source. In: Optical Society of Korea Annual Meeting, pp. 299–300 (2007)
5. Ryu, J.-R., Lee, B.-H.: Design and implementation of high luminous line LED illuminator. J. Adv. Inf. Technol. Convergence **6**(5), 161–169 (2008)
6. Wei, W., Xu, Q., Wang, L., et al.: GI/Geom/1 queue based on communication model for mesh networks. Int. J. Commun Syst **27**(11), 3013–3029 (2014)
7. Song, H., Brandt-Pearce, M.: A 2-D discrete-time model of physical impair-ments in wavelength-division multiplexing systems. J. Lightwave Technol. **30**(5), 713–726 (2012)

An Alternative Management Scheme of DHCP Lease Time for Internet of Things

Pyung Soo Kim(✉), Eung Hyuk Lee, and Eung Tae Kim

System Software Solution Lab., Korea Polytechnic University,
237 Sangidaehak-ro, Siheung-si, Gyeonggi-do 429-793, Korea
pskim@kpu.ac.kr

Abstract. In this paper, an alternative management scheme is proposed for the DHCP lease time in Internet of Things(IoT) environment. In the proposed scheme, the Dynamic Host Configuration Protocol(DHCP) server manages addresses as well as lease time using the type of IoT device. The MAC header of DHCP message sent by IoT device is modified newly by adding specific information to reflect the type of IoT device. The DHCP server can recognize the type of IoT device from the newly defined MAC header and return an address to the IoT device with a lease time determined by the type of IoT device. For the efficient management, a policy for allocating address and lease time should be required. According to this policy, allocated addresses and lease times are stored in a lease database. Examples of the table for diverse IoT types and the policy for allocating address and lease time are given.

Keywords: DHCP · IP address · Lease time · IoT

1 Introduction

Recently, Internet of Things(IoT) is a promising and rapidly developing technology area. IoT has become a powerful force affecting the wide variety of industries, and has been applied in smart home, manufacturing, healthcare, and transportation [1–3].

With the arriving of IoT, the number of devices that are connected to the Internet is growing exponentially. Although IPv4 has been the most dominant internet protocol, the recent exponential growth of Internet-enabled devices and their increasing requirement of IP addresses cannot be fulfilled because of the limited number of IP addresses offered by the IPv4 address space. Thus, the IPv6 is considered the most suitable technology for all the emerging IoT devices, since it spreads the addressing space and offers scalability, flexibility, tested, extended, ubiquitous, open, and end-to-end connectivity [4–6]. However, although the transition to IPv6 is inevitable, it seems to take so much time because IPv4 and IPv6 networks are not directly inter-operable for long time [7]. Therefore, the Dynamic Host Configuration Protocol (DHCP) is still a short-term solution to resolve this problem [8, 9]. The DHCP enables network administrators to automate the assignment of IP addresses from a centralized DHCP server. This conserves IP addresses and make is easy for mobile Internet devices to move among different segments of the network without having to manually enter new IP addresses.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_84

DHCP adopts the concept of a "lease" in IP allocation. This means, a DHCP server does not allocate an IP address to a client permanently. What it does instead is set a "lease time" and allow the DHCP client to use the allocated IP address only during the set lease time. The choice of lease time, like so many other networking parameters, boils down to a trade-off between stability and allocation efficiency. Although many DHCP administrators prefer to use short leases, issuing short lease will result in an overloaded DHCP server and network segments flooded with DHCP traffic. Thus, the seemingly simple and straightforward solution of having the DHCP server only grant short leases cannot be a good solution for the network environment in which Internet devices exceed available addresses.

Therefore, in this paper, an efficient scheme is proposed for the DHCP lease process in IoT environment. In the proposed scheme, the DHCP server manages both addresses and lease time using the type of IoT device. The MAC header of DHCP message sent by IoT device is modified newly by adding specific information to reflect the type of IoT device. Specific information includes the domain type and the device type of diverse IoT devices. The DHCP server receives an address request from the IoT device and then can recognize the type of IoT device from the newly defined MAC header. Then, the DHCP server returns an address to the IoT device with a lease time determined by the type of IoT device. For the efficient management, a policy for allocating address and lease time should be required. Allocated addresses and lease times are stored in a lease database according to this policy. The lease database that includes lease time, MAC addresses, IoT type, etc., can be kept in the DHCP server, or elsewhere on the network for lookup use. For actual implementation, examples of the table for diverse IoT types and the policy for allocating address and lease time are given.

2 Limitation of Current DHCP Lease Time

The DHCP is a network protocol used to assign IP addresses and provide configuration information to devices such as servers, desktops, or mobile devices, so they can communicate on a network. There are four basic phases required in DHCP operations between a DHCP server and DHCP client (e.g. an IoT device) in order for the IoT device to get/lease network configuration data.

DHCP's most significant new feature is dynamic allocation, which changes the way that IP addresses are managed. Where in traditional IP each device owns a particular IP address, in DHCP the server owns all the addresses in the address pool, and each IoT device leases an address from the server, usually only for a limited period of time. That is, like a lease on a car or apartment, a DHCP server does not allocate an IP address to the IoT device permanently. What it does instead is set a "lease time" and allow the IoT device to use the allocated IP address only during the set lease duration. If the IoT device wishes to use the allocated IP address for longer than the lease time, it should request the DHCP server for renewal of the lease. If not, it performs an IP address release procedure instead.

When DHCP dynamic address allocation is used, the administrator of the network must provide parameters to the DHCP server to control how leases are assigned and managed. One of the most important decisions to be made is the lease time policy of

the internetwork: how long the administrator wants IoT device leases to last. There of course is no "right answer"—the right lease time length depends on the network, the server, and the IoT device. The choice of lease time, like so many other networking parameters, boils down to a trade-off between stability and allocation efficiency.

The lease time can be long and even infinite only if the network has a large number of IP addresses available and is in a more stable environment where a configuration rarely changes. On the other hand, the short lease time should be used to free up unused addresses and make them available for the network that have a limited number of IP addresses available, or IoT device configurations that change frequently, or IoT devices that relocate often. Although reducing the lease time increases the rate at which addresses are returned to the available address pool for reassignment, issuing short lease will result in an overloaded DHCP server and network segments flooded with DHCP traffic. Many DHCP administrators prefer to use short leases. This forces an IoT device to continually renew the lease as long as it needs it. When it stops asking for permission again, the address is quickly put back into the pool. This makes shorter leases a better idea in environments where the number of addresses is limited and must be conserved. The drawback, of course, is constantly-changing IP addresses. Thus, the seemingly simple and straightforward solution of having the DHCP server only grant short leases cannot be a good solution for the network environment in which IoT devices exceed available addresses.

3 An Alternative Management Scheme of DHCP Lease Time

As shown in Fig. 1, the DHCP message sent by IoT device includes MAC header. The MAC header is 14 bytes, 6 bytes for the destination MAC address, 6 bytes for the source MAC address, and 2 bytes for the "EtherType". EtherType is used to indicate which protocol is encapsulated in the payload of the MAC header. Usually it is hexa decimal 0800 for IPv4 or 0806 for ARP, but others can be observed sometimes as well (IPv6 coming up with 86DD).

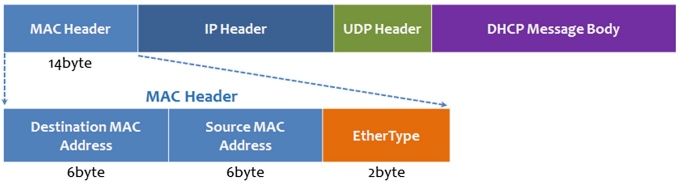


Fig. 1. Ethernet header in DHCP message

In this paper, the MAC header of DHCP message is modified newly by adding specific information to reflect the type of IoT device. Specific information includes the domain type and the device type of diverse IoT devices. That is, to create IoT device specific DHCP message, the EtherType is replaced by 8-bit upper hexadecimal number

for domain type and 8-bit lower hexadecimal number for device type, respectively. As shown Fig. 2, two 8-bit hexadecimal numbers are combined. This 16-bit hexadecimal number is called the "IoT Type" in this paper. Table 1 shows an example of IoT types of diverse IoT devices.

Fig. 2. Ethernet header in DHCP message

Table 1. Example of Diverse IoT Types

Domain type	Upper 8-bit	Device type	Lower 8-bit	Iot type (16-bit)
Mobile	0xF0	Phone	0xF0	0xF0F0
		Tablet	0xF1	0xF0F1
		Laptop	0xF2	0xF0F2
		⋮	⋮	⋮
Sensor	0xF1	Long-time monitoring	0xF0	0xF1F0
		Short-time monitoring	0xF1	0xF1F1
		Always-on	0xF2	0xF1F2
		⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮
Registered thing	0xFF	Employee, Staff, Fixed devices, etc.	0xFF	0xFFFF

The DHCP server receives an address request from the IoT device and then can recognize the type of IoT device from the newly defined MAC header. Then, the DHCP server returns an address to the IoT device with a lease time determined by the IoT type. A key decision is what the network's lease time policy will be. As mentioned, longer leases allow IoT devices to avoid changing addresses too often, while shorter leases are more efficient in terms of reallocating addresses that are no longer required.

For the efficient management in the proposed scheme, a policy for allocating address and lease time should be required. And thus, Table 2 shows an example of the policy when the address pool is set from "192.168.10.1" to "192.168.10.254". Four kinds of lease time are categorized by short, medium, long, and registered according to the lease time length. As an example, highly mobile devices such as smart phones and tablets may be given shorter lease time, while less mobile devices, such as laptops or printers may be given longer lease time. In addition, employees, workers, faculties, doctors in company, institution, hospital, etc., may wish to have longer lease time. Thus, the DHCP administrator can choose from a variety of different lease time, and may choose longer or shorter leases for some IoT devices than for others.

According to this policy, allocated addresses and lease times are stored in a lease database. The lease database can include lease time, MAC addresses, IoT type, etc., and can be kept in the DHCP server, or elsewhere on the network for lookup use. An example of lease database is given in Table 3, which shows the file "`/var/lib/dhcp/dhcpd.leases`" in Red Hat Linux system. For example, a mobile phone with MAC address "00:11:22:33:44:55" and IoT name "phone_1" is given IoT type "0xF0F0", address "192.168.1.1" and lease time "20 min".

Table 2. Example of policy for allocating address and lease time

Classification	Lease time	Range of leasing address (192.168.10.1 ~ 254)	Device example
Short	20 min	128 addresses (192.168.10.1 ~ 128)	Phone, Short-time monitoring sensor
Medium	1 h	64 addresses (192.168.10.129 ~ 192)	Tablet
Long	4 h	32 addresses (192.168.10.193 ~ 224)	Laptop, Long-time monitoring sensor
Registered	8 h	31 addresses (192.168.10.225 ~ 254)	Employee, Staff, Worker, Fixed devices, etc.

Table 3. Example of lease database

```
lease 192.168.1.1 {
  starts 5 2016/06/1 01:57:17;
  ends 5 2016/06/1 02:17:17;
  tstp 5 2016/06/1 02:17:17;
  cltt 5 2016/06/1 01:57:17;
  binding state free;
  hardware ethernet 00:11:22:33:44:55;
  IoT name "phone_1";
  IoT type 0xF0F0;
}
lease 192.168.1.2 {
  starts 5 2016/06/1 02:47:17;
  ends 5 2016/06/1 03:07:17;
  tstp 5 2016/06/1 03:07:17;
  cltt 5 2016/06/1 02:47:17;
  binding state free;
  hardware ethernet 00:11:22:33:44:66;
  IoT name "phone_2";
  IoT type 0xF0F0;
}
lease 192.168.1.127 {
  starts 5 2016/06/1 02:27:17;
  ends 5 2016/06/1 03:27:17;
  tstp 5 2016/06/1 03:27:17;
  cltt 5 2016/06/1 02:27:17;
  binding state free;
  hardware ethernet 00:22:33:44:55:66;
  IoT name "tablet_1";
  IoT type 0xF0F1;
}
lease 192.168.1.192 {
  starts 5 2016/06/1 01:27:17;
  ends 5 2016/06/1 05:27:17;
  tstp 5 2016/06/1 05:27:17;
  cltt 5 2016/06/1 01:27:17;
  binding state free;
  hardware ethernet 00:33:11:22:55:77;
  IoT name "labtop_1";
  IoT type 0xF0F2;
}
lease 192.168.1.230{
  starts 5 2016/06/1 09:00:00;
  ends 5 2016/06/1 17:00:00;
  tstp 5 2016/06/1 17:00:00;
  cltt 5 2016/06/1 09:00:00;
  binding state free;
  hardware ethernet 00:55:11:22:66:99;
  IoT name "doctor_1";
  IoT type 0xFFFF;
  }
```

4 Conclusion

This paper has proposed an alternative management scheme for the DHCP lease time in IoT environment. In the proposed scheme, the DHCP server has managed both addresses and lease time using the type of IoT device. The MAC header of DHCP message sent by IoT device has been modified newly by adding specific information to reflect the type of IoT device. The DHCP server can recognize the type of IoT device from the newly defined MAC header and return an address to the IoT device with a lease time determined by the type of IoT device. For the efficient management, a policy for allocating address and lease time should be required and thus has been given. According to this policy, allocated addresses and lease times are stored in a lease database. Examples of the table for diverse IoT types and the policy for allocating address and lease time have been given.

As a future work, the policy for allocating address and lease time should be extended with reflecting other properties of IoT devices such as location, time of day, application, and traffic amount.

Acknowledgment. This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the ITRC(Information Technology Research Center) support program (IITP–2016-H8601–16–1003) supervised by the IITP(Institute for Information & communications Technology Promotion).

References

1. Palattella, M.R., Accettura, N., Vilajosana, X., Watteyne, T., Grieco, L.A., Boggia, G., Dohler, M.: Standardized protocol stack for the Internet of (Important) Things. IEEE Commun. Surveys Tutorials **15**(3), 1389–1406 (2013)
2. Perera, A., Zaslavsky, P.C., Georgakopoulos, D.: Context aware computing for the Internet of Things: a survey. IEEE Commun. Surveys Tutorials **16**(1), 414–454 (2014)
3. Xu, D., He, W., Li, S.: Internet of Things in industries: a survey. IEEE Trans. Ind. Inform. **10**(4), 2233–2243 (2014)
4. Roberts, P., Ford, M.: World IPv6 launch: the future is forever. IETF J. **8**(1), 12 (2012)
5. Jara, A.J., Ladid, L., Skarmeta, A.: The Internet of Everything through IPv6: an analysis of challenges, solutions and opportunities. J. Wirel. Mob. Netw. Ubiquitous Comput. Dependable Appl. (JoWUA) **4**(3), 97–118 (2013)
6. Savolainen, T., Soininen, J., Silverajan, B.: IPv6 addressing strategies for IoT. IEEE Sens. J. **13**(10), 3511–3519 (2013)
7. Kim, P.S.: Comparison and analysis of IPv4/IPv6 transition technologies. Telecommun. Rev. **24**(3), 419–432 (2014)
8. Droms, R.: Dynamic Host Configuration Protocol, IETF RFC 2131, March 1997
9. Kozierok, C.M.: The TCP/IP Guide: A Comprehensive, Illustrated Internet Protocols Reference. No Starch Press (2005)

A User Empirical Context Model for a Smart Home Simulator

Green Bang and Ilju Ko(✉)

Department of Information Communication, Materials, and Chemistry Convergence Technology, Soongsil University,
369 Sangdoro, Dongjak-Gu, Seoul, Korea
{banglgreen, andyl026}@gmail.com

Abstract. This dissertation aims to quantify the consumer experience in reality by building a real-space information based space model and consumer behavior model. We applied the models and used it to implement the simulator. The developed simulator collects the information generated in the virtual living space for the purpose of developing a smart home simulator. To predict user behavior, data is compiled from experiment panels. By using weka, features are extracted from the amassed data and the feature value's complexity is reduced. Through optimal subsets of the feature values we could make a judgment on the accuracy of the predicted behaviors.

Keywords: Smart home · User experience · User behavior · Simulator

1 Introduction

As the social paradigm shifts to allow for more anthropocentric values, the time is ripe for a society in which it is crucial to offer the user experience optimized for his or her life style or situation. User experience can be defined as an overall experience that a user goes through while using a system, a product, or a service [1]. In addition, as interest escalates for the physical space in which the user's experience takes place, so does the interest for the user's proactive experience of space which adjusts accordingly to changes in properties of the said space. On the arrival of the epoch of the Internet of Everything, in which every object and its respective data set is connected real time, the need arises for the host of user related data to be controlled and processed in order to be utilized efficiently [2, 3]. In order to make use of consumer related experience, quantifying is essential and for that the user empirical context model is necessary. The user empirical context model refers to the model that quantifies and models all of the interdependent circle of information of space, object, and people that surrounds a person. In this paper, we aim to make use of the simulator that is loaded with the user empirical context model to collect user behavior information, which is otherwise difficult to do due to constraints in budgets, space, and legality, and conduct research in predicting user behavior.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_85

2 Related Works

We plan to examine the methodology of simulator development and the modeling technique for quantifying the user empirical information by looking at the related research. UbiREAL is the simulator designed to simulate the interaction system composition between man and device in a context-based smart environment through the illustrated monitor [5]. The sensors arranged with the objective of situational awareness perceives the character that moves according to the designated path and generates the data. The system is designed to recognize the rules in the interaction between the multitude of users and devices in the target space and discern what changes occur in the device's movement in relation to changes in the context. This simulator deserves some merit as it is meant to create a networked simulator by utilizing a ubiquitous application. But the simulator fails to escape the experimental stage and could only produce meaningful results in a restricted physical capacity. Furthermore, active experiment in real life is challenging because the error rate in reality was not taken into account.

Persim 3D is the simulator that makes use of virtual sensors in virtual space to perceive movement of the character and thus figures out the user's moving patterns in the smart home [6, 7]. However, the smart character in this simulator cannot be operated manually but rather operates passively following a preset scenario; it does not contain the user's behavior information. For the character to operate more proactively in relation to the user's behavior, and to fully realize the correlation between the space-object-user, a fragmented dataset of the target environment is not sufficient. We could also find other smart home simulators such as SIMACT and CASAS [8–11].

3 Smart Home Simulator Realization and Result Analysis

To accurately predict the user's behavior, sufficient amount of data for each type of behavior is essential and the simulator should be constructed so that there is no difficulty with the data collection. The simulator loaded with user empirical context model was modeled based on data of real users and objects. The user empirical context model comprises the user behavior model and the space model. User behavior was modeled based on real space user behavior patterns and the space model was constructed by using space and object information collected by sensors. The space model provides information that is needed to map out virtual space interspace and Table 1 below is an

Table 1. Application of the space model in the residential space

Partial space	Partial zone		
	Specialized zone	Retention zone	Convenience zone
Living room	TV	Sofa	Light Computer
Kitchen	Gas cooker Sink Refrigerator	Table Chair	Shelves Light

example of application of the space model on a residential space. The residential space is divided into the living room, kitchen, entrance, bathroom, and bedroom. In each space, ordinary everyday objects were arranged. Each user can choose to add or remove objects. The example shows the finer details of the living room and kitchen and the related objects (Table 2).

And Fig. 1(a) is a developed smart home simulator use Unity 3D engine for Game. In the implemented space, virtual sensors exist. Camera sensor, pressure sensor, touch sensor, temperature sensor, human detection sensor and other sensors recognize user's behavior and whether an object is being used. And (b) is an example about recognizing of object in the virtual space. As one of the capabilities provided by Unity 3D. We can find the object in virtual residential space about character's behavior.

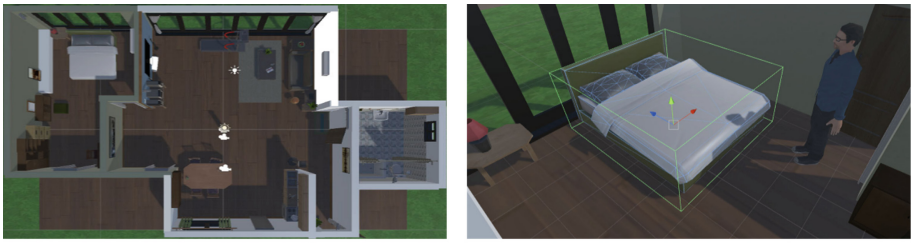


Fig. 1. (a) is a developed Smart home simulator and (b) is an example about recognizing of object I the virtual space.

Table 2. Type of the behavior model in the residential space

Behavior type	Behavior value
Preparing meals and cleaning up	Preparing meal
	Clean up after meal
Food and drink related	Taking a meal
	Eating snacks
Cleaning and organizing	Cleaning
	Organizing
Sleep	Sleep
	Nap
Personal care	Changing clothes
	Putting on make-up
Personal sanitation	Washing
	Defecation, Urination

A key factor that must be considered when designing an user behavior model is that a clear standard for classifying actions is difficult to suggest. Although there is the problem of vague standards when defining an action, to increase the accuracy of the

Table 3. User behavior feature classification

Classification	Feature
Location	Current location in space
	Current location in particular area
Action	Previous action information
Object	Object ID being used
	Whether the object is used
Time	Time of event occurrence
	Time of event sustainment
Environment information	Temperature
	Weather
	Day of week

behavior prediction a feature selecting experiment was done to categorize the variable activities. Table 3 is a categorization of the user's activity features and the feature values that were used to extract the features.

The features of the data collected by the simulator was extracted by Weka, an open software that made the Support Vector Machine algorithm easy to use. In SVM the most optimal subset of the feature values were found to decrease the complexity of the feature values and increase efficiency when predicting user behavior. Table 4 is a rank result table of the optimal subsets of the features gathered by a simulator with the user empirical environment model loaded. It displays the components of the subsets ranking from first to fourth. (N = 6,853).

Table 4. Feature extraction experiment result

Ranking	Point	Features subset
1	0.96369	Object ID being used, Whether the object is used, Current location in space, Time of event occurrence, Previous action information
2	0.96165	Object ID being used, Whether the object is used, Current location in space, Time of event occurrence Weather, Previous action information
3	0.96042	Object ID being used, Whether the object is used, Time of event occurrence, Previous action information

Table 5 is a comparison of accuracy when predicting user behavior of the 3 experiment subjects.

The user behavior prediction result indicates that the activities over the accuracy of 90 % in sample 1, 2, 3 was preparing meal, cleaning, reading, clean up after meals, and watching TV starting from the highest accuracy in consecutive order. However, in the case of rest the prediction accuracy was very low. Rest's definition and behavior depends on the user so an understanding of this is needed.

Table 5. Accuracy of user behavior prediction

Behavior value	Accuracy		
	Sample 1	Sample 2	Sample 3
Preparing meal	0.982	0.967	0.983
Clean up after meals	0.957	0.912	0.932
Snacks	0.938	0.845	0.924
Cleaning	0.979	0.964	0.973
Cleaning products	0.904	0.887	0.902
Wash	0.894	0.895	0.894
Using the toilet	0.791	0.935	0.877
Dressing change	0.894	0.845	0.872
Make up	0.871	0.937	0.912
Organize clothes	0.758	0.875	0.831
Reading	0.972	0.967	0.985
Watching TV	0.892	0.942	0.933
Using a computer	0.824	0.938	0.942
Rest	0.780	0.804	0.743

4 Conclusion

In this thesis we have strived to create a basis to quantify the user experience information of a highly complicated space in reality. Using a simulator loaded with a user experience environment model, user behavior information is gathered and the user's actions are predicted through learning. With further studies the user experience environment model will be extended and there will be efforts to create a frame that will be applicable to various spaces. Studies like this is expected to be applicable in evaluation user experience.

References

1. Hartson, R., Pyla, P.: The UX Book: Process and Guidelines for Ensuring a Quality User Experience. Morgan Kaufmann, San Diego (2012)
2. Garrett, J.J.: The Elements of User Experience: User-Centered Design for the Web and Beyond, 2 edn. New Riders Publishing, Berkeley (2003)
3. Cisco: Internet of Everything (IoE). Cisco and/or its affiliates (2013)
4. Evans, D.: The Internet of Everything: How More Relevant and Valuable Connections Will Change the World. 2012 Cisco and/or its affiliates, pp. 1–8 (2012)
5. Nishikawa, H., Yamamoto, S., Tamai, M., Nishigaki, K., Kitani, T., Shibata, N., Yasumoto, K., Ito, M.: UbiREAL: realistic smartspace simulator for systematic testing. In: Dourish, P., Friday, A. (eds.) UbiComp 2006. LNCS, vol. 4206, pp. 459–476. Springer, Heidelberg (2006). doi:10.1007/11853565_27

6. Elfaham, A., Hagras, H., Helal, S., Hossain, S., Lee, J.W., Cook, D.J.: A fuzzy based verification agent for persim - an ambient intelligent environment simulator. In: 2010 IEEE International Conference on Fuzzy Systems (FUZZ), pp. 1098–7584 (2010)
7. Helal, S., Lee, J.W., Hossain, S., Kim, E., Hagras, H., Cook, D.J.: Persim – simulator for human activities in pervasive spaces. In: Proceedings of the 7th International Conference on Intelligent Environments, pp. 192–199 (2011)
8. Cook, D.J., Schmitter-edgecombe, M., Crandall, A.S., Sanders, C., Thomas, B.L.: Collecting and disseminating smart home sensor data in the CASAS project. In: Proceedings of the CHI Workshop on Developing Shared Home Behavior Datasets to Advance HCI and Ubiquitous Computing Research, pp. 1–9 (2009)
9. Bouchard, K., Ajroud, A., Bouchard, B., Bouzouane, A.: SIMACT: a 3D open source smart home simulator for activity recognition with open database and visual editor. Intl. J. Hybrid Inf. Technol. **5**(3), 13–32 (2012)
10. Buchmayr, M., Kurschl, W., Küng, J.: A simulator for generating and visualizing sensor data for ambient intelligence environments. Procedia Comput. Sci. **5**, 90–97 (2011)
11. Kim, E., Helal, S., Lee, J., Hossain, S.: The making of a dataset for smart spaces. In: Yu, Z., Liscano, R., Chen, G., Zhang, D., Zhou, X. (eds.) UIC 2010. LNCS, vol. 6406, pp. 110–124. Springer, Heidelberg (2010). doi:10.1007/978-3-642-16355-5_11

Co-display Content Service for First-Person Videos of Smart Glass

Bokyung Sung[1] and Ilju Ko[2(✉)]

[1] PDK LIMITED Co., 369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
bksung@pdklimited.com
[2] Department of Information Communication, Materials, and Chemistry Convergence Technology, Soongsil University,
369 Sangdo-ro, Dongjak-gu, Seoul 06978, Korea
andyl026@gmail.com

Abstract. A smart glass is a personalized device, which has features that enable it to record and share a user's experience of specific events in first-person view (FPV). FPV videos recorded from identical events can be watched interactively, for example, it is possible to switch between FPV videos of several people through synchronization based on an actual event timeline. This paper proposes a co-display content service for watching interactive, synchronized FPV videos in second screen environments and presents a prototype of the system.

Keywords: Smart glass · First-person view (FPV) · Synchronized FPV video · Co-display · Content service

1 Introduction

Recently, various types of wearable devices that can connect to smartphones and smart TVs have been launched. Wearable devices show the possibility of increasing smartphone usability and even replacing the smartphone itself. Unlike 3D glass and head-mounted display (HMD), which function simply as additional devices for watching videos, the glass type of device with recording function is able to input and output data. That is, wearable glass type devices can both produce and consume content. Objects observed from the user's perspective can also be recorded from that vantage point via a camera function operating around the eyes—the closest place for a first-person perspective. This facilitates the recording and sharing of first-person view (FPV) experiences [1, 2]. It also makes it possible to offer and consume information in real time and view the angle of the user through a display function. Methods for servicing various content types based on the features outlined have been proposed.

Smart glass is fundamentally a personalized device and the camera function is the one that is primarily used. As a result, it is being predicted that active personal recording videos will be one kind of user-created content produced with a view towards documentation [3]. It can also be considered an exclusive service of smart glass's features to be able to record directly from a user's viewpoint.

Videos that are recorded from a personal viewpoint can be used to document personal experiences and the recording and sharing of social activities such as a

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_86

performance or a party attended with friends and family. There is also a tendency for users to not only watch recorded video together with people but to also be active in the same event, which is preferred more than watching videos alone [4]. Further, the playback or live streaming of the recording is preferred on a wide display or smartphone rather than the actual device. Within this context, methods for playing the video that is recorded with several people together through co-display devices such as TVs are presented.

A collection of FPV videos obtained from several people at the same event, where each person provides a unique view of various attendees at the event, amounts to multi-view video content of an event. To apply multi-view features, FPV videos recorded from an event can be watched by switching among the views of the event attendees. One problem in this scenario is that of aligning the time axis among the FPV videos. The main purpose of watching recorded videos is to reproduce past experiences. FPV videos that include attendees' views and are recorded at identical moments should be presented at the moment the attendee's view switches because the FPV videos of attendees are from one group. To facilitate this, all FPV videos from one group should be synchronized on the timeline of the actual event. Differentiation of the time length of the FPV videos and the number of each attendee's FPV video is a feature of user-created videos recorded by people attending the same event.

Applying audio signals, instead of video frames, is recommended for synchronization between FPV videos that are recorded with views of each attendee at the identical event, because although the videos are recorded of the same event, there is a high possibility that not every attendee will record the same object by an identical view. That is, with video frames, it is difficult to guarantee constant information flow for synchronization owing to diverse attendees' views. In contrast, audio content is identical, excluding factors such as environment noise and volume levels. For this reason, requested information amount for synchronization should be maintained, regardless of the changing view of the person who is recording.

2 Acquiring and Synchronizing FPV Videos via Smart Glass

In this paper, the precondition for FPV videos is that several people take videos while having identical experiences via their own smart glass. Figure 1 illustrates the precondition: Three attendees at the same music concert enjoy and record the event in FPV from a different location and view angle. That is, FPV videos that are recorded in the event by more than one person who get together are a basis unit for synchronization.

The events under consideration include not only music concert but also ordinary events that can be co-experienced by several people at the same time. Representative examples of these events include "music concerts," "regular parades," "birthday parties," "anniversaries," "social meetings," "museum exhibitions," "formal dinners," "weddings," and "political demonstrations" [5]. Each event presents attributes that differentiate them such as accruing periodicity from single to recurring and social type from public to private; however, they have the common feature that co-experience occurs within their ambit. However, FPV videos are not made at every event. Three conditions—space, time, and relation—must all be satisfied at the same time. First, there must be at

least two persons who adjudge that the attended event is significant and so they should record it. Second, attendees should be in a limited place that is recognized as one identical event. Third, attendees should be in the same limited timeline that is recognized as one identical event.

FPV videos that satisfy these conditions present features the start and end times of FPV videos should be different. The length of FPV videos differ from user to user; however, an overlap in which a common time zone exist is shared among the FPV videos of users. There is a strong possibility that the parts of each FPV video that overlap do not include the same object because of differing view angles of the users.

However, FPV videos that include overlapping parts that occur at the same event but have a different image that is not appropriate for judging them identical can be included in the FPV videos. Conversely, all the audio of the overlapping part of each FPV video are similar even when noise such as sounds from the environment is included. For this reason, FPV videos were synchronized on the basis of the audio signal.

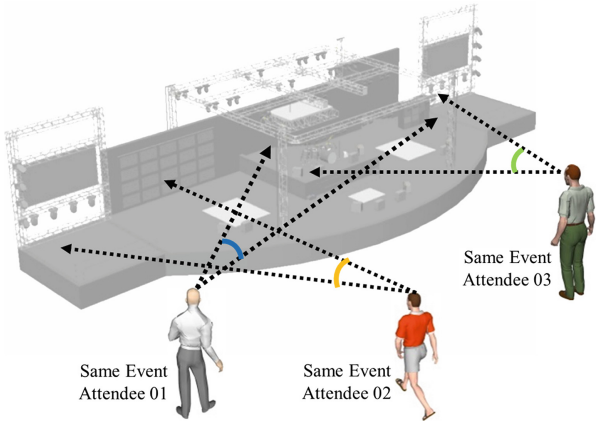


Fig. 1. Example of ideal environment for taking FPV videos at same event

The audio of FPV videos is transformed into an appropriate form for extracting feature data by means of preprocessing. At the preprocessing stage, different channel numbers, sound quality, format, and waveform are normalized to the same condition. Minimal sound quality formatting was applied to videos distributed online so as not to affect sound recognition. The format used was mono channel, 64 kbps bit rate and 22050 Hz sampling.

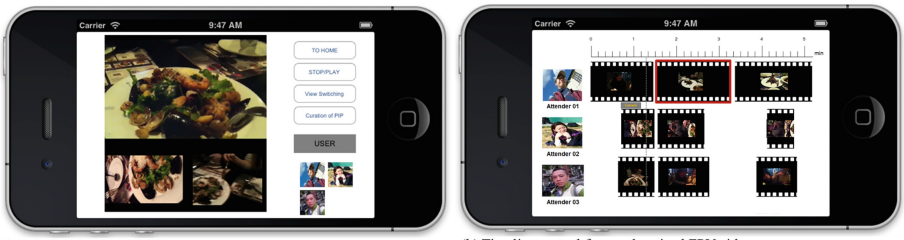
We applied simplified-delta (SD) Mel Frequency Cepstrum Coefficients (MFCCs) [6], which are generally used to extract features from preprocessed audio waves. SD-MFFCCs apply simplified-delta to extracted feature sequences using MFCCs for simplification of extracted feature data.

After undergoing the similarity comparison method, synchronized FPV video can be watched by switching the view point of event attendees. A mobile application including an interface for co-display of classified FPV videos of event attendees was

implemented. The application comprised a playing interface for co-display of synchronized FPV videos and a control interface for switching views and changing the timeline while playing synchronized FPV video, as depicted in Fig. 2.

The co-display interface consists of an FPV videos display area, a function area for connecting with the control interface, and an attendee information area. In the FPV videos display area, several FPV videos that were recorded in the same timeline are presented. This area is divided into one main area for displaying the selected FPV video and two sub-areas for displaying the remaining FPV videos (Fig. 2(a)).

Playing and stopping of synchronized FPV videos signify playing and stop of one synchronized video along the actual event timeline. Figure 2(b) shows that synchronized status in which attendees and their FPV video clips of synchronized FPV video are arranged by one timeline. That is, users can put the attendee's view that the user wants into the main area or change the time spot that the user wants by selecting a particular FPV video clip or time spot.



(a) Playing screen of synchronized FPV video

(b) Timeline control for synchronized FPV video

Fig. 2. Mobile application for auto playing synchronized FPV video with timeline control function

3 Co-display Content Service for Synchronized FPV Videos

In this paper, the purpose of the co-display content service is to present a prototype based on home media network of the Digital living Network Alliance (DLNA) [7]. The representative device category of DLNA for media share and control on home network is divided into Home Network Device (HND) and Mobile Handheld Device (MHD). Then, each category has its own class definition. HND comprises components such as Digital Media Server (DMS), Digital Media Player (DMP), Digital Media Renderer (DMR), and Digital Media Controller (DMC). MHD comprises components such as Mobile Digital Media Uploader (M-DMU), Mobile Digital Media Server (M-DMS), Mobile Digital Media Player (M-DMP), and Mobile Digital Media Controller (M-DMC).

Figure 3 shows the architecture of the proposed co-display content service for synchronizing and playing FPV videos. M-DMU plays the role of media generator; that is, video is generated by devices that have a camera function included. A smart glass is suitable as an M-DMU because FPV videos are being used. DMS shares the media: it collects videos from the M-DMUs and sends them to devices for playback. The DMS function could be a service from a wireless carrier aspect like the cloud or software

from a device manufacturer aspect like local applications preinstalled on a device. The DMR included in the range of DMS plays the role of media retrieval; that is, the DMR conducts synchronization during DMS processing. The DMP carries out media reproduction; that is, FPV video synchronized by the DMR is sent to the DMP through the DMS and is then reproduced along the control message of the M-DMC. Smart TV is suitable for giving public displays in the home. The M-DMC carries out event retrieval and control for media reproduction; that is, the user selects or retrieves FPV video clips in synchronized FPV video using the control interface UI. Further, user control switches view and timeline through the control interface UI. A smart pad is suitable as an additional tool for richer control than a remote controller of smart TVs.

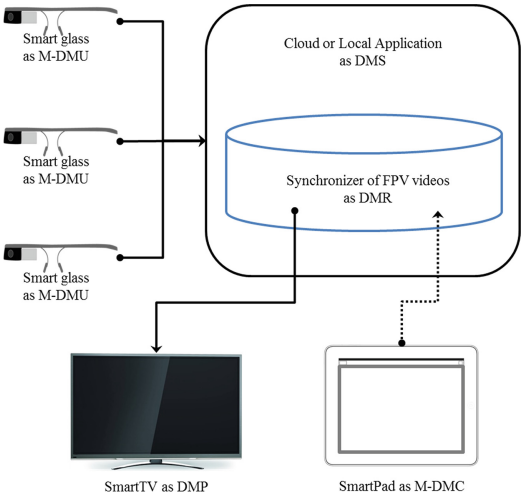
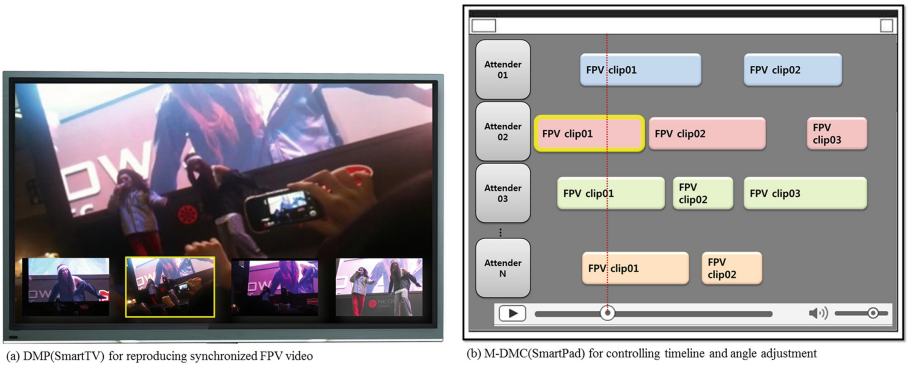


Fig. 3. Architecture of proposed co-display content service



(a) DMP(SmartTV) for reproducing synchronized FPV video

(b) M-DMC(SmartPad) for controlling timeline and angle adjustment

Fig. 4. Prototype for the application of synchronized FPV video to second screen service of co-display form

FPV videos can be composed as a co-display content service in the DLNA environment and consumed in co-display form. Figure 4 shows a sample interface of DMP and M-DMC in the prototype of the co-display content service. The lower part of Fig. 4 shows the structure of the synchronized FPV video for the criteria of single common timeline and M-DMC interface for playing control by switching the attendee's view. The upper part of Fig. 4 shows an interface for playing synchronized FPV videos in the DMP by controlling the M-DMC.

4 Conclusions

This paper proposes a co-display method that facilitates the watching of FPV videos created from several FPV videos shot by event attendees with attendees' views synchronized and switchable. A device similar to a smart glass that can shoot videos near the user's eyes to take FPV videos was used. FPV videos with similar timelines and past actual event timelines were arranged by recognizing the overlapping parts based on audio matching for synchronization between FPV videos. The synchronized FPV video includes attendees' FPV videos and a reproduced timeline. Thus, users can watch existing events from various viewpoints simply by switching the attendees' view. A prototype of the DLNA environment for synchronized FPV video was presented because playback on a second screen, through which one can watch and control the video at the same time, is suitable for this kind of video.

In the future, the volunteers will be experimented for the usability valuation survey.

Acknowledgements. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (No. 2012M3C4A7032783)

References

1. Ryoo, M.S., Matthies, L.: First-person activity recognition: what are they doing to me. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2730–2737 (2013)
2. Fathi, A., Hodgins, J.K., Rehg, J.M.: Social interactions: a first-person perspective. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1226–1233 (2012)
3. Starner, T.: Project glass: an extension of the self. Pervasive Comput. IEEE **12**(2), 14–16 (2013)
4. Clinch, S.: Smartphones and pervasive public displays. Pervasive Comput. IEEE **12**(1), 92–95 (2013)
5. Kaplun, V., et al.: Understanding media capture and consumption for live music event. Technical report, Yahoo!Inc (2008)
6. Logan, B.: Mel frequency cepstral coefficients for music modeling. In: Proceeding of ISMIR International Symposium on Music Information Retrieval (2000)
7. Scott, S.: Digital home standards: choosing and implementing the right ones. In: IEEE DLNA presentation (2007)

Probabilistic Analysis for the Relationship Between Min-Entropy and Guessing Attack

Ju-Sung Kang(✉), Hojoong Park, and Yongjin Yeom

Department of Mathematics and Financial Information Security,
Kookmin University, 77 Jeongneung-Ro, Seongbuk-Gu, Seoul 02707, Korea
{jskang, ruokay, salt}@kookmin.ac.kr

Abstract. Recently NIST has published the second draft document of recommendation for the entropy sources used for random bit generation. In this document NIST has provided a practical and detailed description about the fact that the min-entropy is closely related to the optimum guessing attack cost. However the argument lacks the mathematical rigour. In this paper we provide an elaborate probabilistic analysis for the relationship between the min-entropy and cost of optimum guessing attack. Moreover we also provide some simulation results in order to investigate the practicality of optimum guessing attack.

Keywords: Entropy source · Min-Entropy · Optimum guessing attack

1 Introduction

Entropy is a measure of information or uncertainty, and is computed as a function of a probability distribution. There are many possible types of entropy in addition to Shannon entropy. The min-entropy is the smallest entropy measure in the family of Rényi entropies and has important applications for randomness extractors in cryptography. Especially it provides lower bounds on the required key sizes in some attack scenarios.

Random numbers are crucial to modern cryptography and many information security applications. They are used as confidential keys, nonces and salts in some protocols, initialization vectors, padding values, mask values etc. The notion of min-entropy can be available to test the randomness collected by an operating system or some applications for use in cryptography. A lack of entropy can have a negative impact on performance and security. For this reason, recently NIST has published the second draft document of recommendation for the entropy sources that contains a noise source such as thermal noise or hard drive seek times, used for random bit generation [3]. This document includes entropy source development requirements and tests for determining entropy provided by entropy sources. In order to measure the entropy of an entropy source, NIST uses the min-entropy known as a very conservative measure on the ground that the min-entropy measures the difficulty of guessing the most likely output of the entropy source.

The study on the problem of guessing a secret value was initiated by Massey [2]. He has obtained a lower bound for the average number of successive guesses until one

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_87

correctly guesses with an optimum strategy. Cachin [1] has shown that the min-entropy is closely related to the optimum guessing attack cost, and NIST [3] also has provided more practical and detailed description about this relationship. It seems that the description in [3] is a convincing argument to claim the legitimacy for the usage of min-entropy as a measure of cryptographic entropy sources but for the argument lacks the mathematical rigour. In this paper we provide a rigorous probabilistic analysis for the relationship between the min-entropy and cost of optimum guessing attack. Our result contributes to solidify the theoretical foundation of the relationship. We also provide some simulation results in order to investigate practicality of optimum guessing attack.

2 Optimum Guessing Attack and Min-Entropy

Consider the problem of guessing the value of a discrete random variable X on $\{x_1, x_2, \ldots, x_m\}$ with probability distribution $\mathbf{P} = (p_1, p_2, \ldots, p_m)$, where $p_i = \mathrm{P}(X = x_i)$, $1 \leq \mathrm{i} \leq \mathrm{m}$. Without loss of generality we assume that $p_1 \geq p_2 \geq \ldots \geq p_m$. Then the min-entropy of X is defined as

$$H_\infty(X) = \min_{1 \leq i \leq m} (-\log_2 p_i) = -\log_2 \max_{1 \leq i \leq m} p_i = -\log_2 p_1.$$

Suppose that an adversary wants to detect at least one of several secret values, where each secret value is independently chosen from the set $\{x_1, x_2, \ldots, x_m\}$ with the identical distribution $\mathbf{P}$. That is, for any $j \geq 1$, let X_j be the random variable of j-th secret value, then $X_1, X_2, \ldots$ are independent and identically distributed random variables with the probability distribution $\mathbf{P}$.

Now we describe the optimum guessing attack scenario with a guessing strategy aimed at successful recoveries as many secret values as possible. The goal of adversary is to minimize the expected number of guesses per successful recovery. We consider the following attack scenario consists of guessing a maximum of $k (1 \leq k \leq m)$ possibilities for a given secret value:

1. For each j-th secret value $X_j (j = 1, 2, \ldots)$, the adversary guesses the realization of X_j by asking questions "Is X_j equal to x_i?" for $1 \leq i \leq k$ in consecutive order, since $p_1 \geq p_2 \geq \ldots \geq p_k$. We call this step as the j-th attack.
2. The adversary moves on to the next attack with a new secret value when either a guess is correct or k incorrect guesses for the current attack have been made.

NIST [3] showed that the min-entropy $H_\infty(X)$ corresponds to the expected work per success of the above attack or over-estimates this work by at most one bit. However the explanation of [3] for the derivation process of the expected work per success is somewhat untransparent. We would like to clarify this process by an elaborate probabilistic approach in the following section.

3 Probabilistic Analysis of Attack Cost

In this section we want to obtain the expected value of total number of guesses per successful recovery in the optimum guessing attack described in the previous section. At first we fix the maximum number of guesses k of the first step in this attack scenario. Let $G_j^{(k)} (j = 1, 2, \ldots)$ be the number of guesses on the j-th attack, $T^{(k)}$ be the number of attack times up to the first successful recovery for a secret value, and $S^{(k)}$ be the total number of guesses up to the first successful recovery. Then $G_j^{(k)}$'s, $T^{(k)}$, and $S^{(k)}$ are random variables depending on the probability distribution $\mathbf{P}$, and we obtain the formula

$$S^{(k)} = \left(T^{(k)} - 1\right)k + G_{T^k}^{(k)}.$$

Lemma 1. For any fixed $1 \le k \le m$, $G_1^{(k)}$, $G_2^{(k)}$, … are a sequence of independent and identically distributed random variables such that for each $j = 1, 2, \ldots$,

$$\mathrm{P}\left(G_j^{(k)} = i\right) = p_i, \text{for } 1 \le i \le k-1, \textit{ and } P\left(G_j^{(k)} = k\right) = 1 - \sum_{i=1}^{k-1} p_i.$$

Proof. By the second step of the attack scenario in the previous section, $G_1^{(k)}$, $G_2^{(k)}$, … are independent, since each attack is moved to the next attack regardless of the current attack has been successful. Note that for any j-th attack, $j = 1, 2, \ldots$,

$$\mathrm{P}\left(G_j^{(k)} = i\right) = \mathrm{P}\left(X_j = x_i\right) = p_i, \textit{for } 1 \le i \le k-1,$$

and

$$\mathrm{P}\left(G_j^{(k)} = k\right) = \mathrm{P}\left(X_j \in \{x_1, x_2, \ldots, x_{k-1}\}\right) = 1 - \sum_{i=1}^{k-1} p_i.$$

This completes the argument. □

Lemma 2. For any fixed $1 \le k \le m$, $T^{(k)}$ is a geometric random variable such that

$$\mathrm{P}\left(T^k = l\right) = \left(1 - \sum_{i=1}^{k} p_i\right)^{l-1} \cdot \sum_{i=1}^{k} p_i, \text{ and } \mathrm{E}\left[T^{(k)}\right] = \frac{1}{\sum_{i=1}^{k} p_i}.$$

Proof. Since attacks are independent and $T^{(k)}$ be the number of attack times until the first successful recovery occurs, it is easy to see that $T^{(k)}$ is a geometric random variable where, for each j-th attack, the success probability is

$$\mathrm{P}(j-th\ attack\ is\ successful) = \mathrm{P}\left(X_j \in \{x_1, x_2, \ldots, x_k\}\right) = \sum_{i=1}^{k} p_i,$$

And it is a well-known fact that the expected value of a geometric random variable is the reciprocal of the success probability. □

Theorem 1. For any fixed $1 \le k \le m$,

$$\mathrm{E}\left[S^{(k)}\right] = \frac{\sum_{i=1}^{k-1} i \cdot p_i + k\left(1 - \sum_{i=1}^{k-1} p_i\right)}{\sum_{i=1}^{k} p_i}.$$

Proof. Recall that, for any fixed $1 \le k \le m$, $S^{(k)} = \left(T^{(k)} - 1\right)k + G^{(k)}_{T^{(k)}}$, then

$$\mathrm{E}\left[S^{(k)}\right] = k\left(E\left[T^{(k)}\right] - 1\right) + E\left[G^{(k)}_{T^{(k)}}\right].$$

It is sufficient to calculate the value of $E\left[G^{(k)}_{T^{(k)}}\right]$, since $E\left[T^{(k)}\right]$ is already obtained in Lemma 2.

$$E\left[G^{(k)}_{T^{(k)}}\right] = E\left[E\left[G^{(k)}_{T^{(k)}} | T^{(k)}\right]\right] = \sum_{l=1}^{\infty} P(T = l) \cdot E\left[G^{(k)}_{l} | T^{(k)} = l\right].$$

On the other hand, for any $1 \le j \le k$ and $l \ge 1$, by Lemma 1,

$$\begin{aligned} P\left(G^{(k)}_{l} = j | T^{(k)} = l\right) &= P\left(G^{(k)}_{l} = j | l - th\ attack\ is\ successful\right) \\ &= \frac{P\left(G^{(k)}_{l} = j\ and\ l - th\ attack\ is\ successful\right)}{P(l - th\ attack\ is\ successful)} \\ &= \frac{p_j}{\sum_{i=1}^{k} p_i}, \end{aligned}$$

and

$$\mathrm{E}\left[G^{(k)}_{l} | T^{(k)} = l\right] = \sum_{j=1}^{k} j \cdot P\left(G^{(k)}_{l} = j | T^{(k)} = l\right) = \frac{\sum_{j=1}^{k} j \cdot p_j}{\sum_{i=1}^{k} p_i}.$$

Thus we obtain that

$$E\left[G^{(k)}_{T^{(k)}}\right] = \frac{\sum_{j=1}^{k} j \cdot p_j}{\sum_{i=1}^{k} p_i} \cdot \sum_{l=1}^{\infty} P(T = l) = \frac{\sum_{j=1}^{k} j \cdot p_j}{\sum_{i=1}^{k} p_i}.$$

Table 1. Estimations for min-entropy and expected cost of guessing attack.

Data set	Min-entropy	$\min_{1 \le k \le m} \log_2 E\left[S^{(k)}\right]$
4-bit data set from [4]	3.9711	3.1457
8-bit data set from [5]	7.9124	6.9004

Consequently the argument is completed as follows:

$$\mathrm{E}\left[S^{(k)}\right] = k\left(\frac{1}{\sum_{i=1}^{k} p_i} - 1\right) + \frac{\sum_{j=1}^{k} j \cdot p_j}{\sum_{i=1}^{k} p_i}. \qquad \square$$

The results of Theorem 1 and [3] come to the same thing. That is, we have proved that $\mathrm{E}\left[S^{(k)}\right] = W_k(\mathbf{P})$ by the rigorous process, where $W_k(\mathbf{P})$ is the expected work per success [3].

4 Experimental Results

We have selected 4-bit data set from truerand_4bit [4] provide by NIST and 8-bit data set from TrueRNG2 [5] of ubld.iT as the experimental targets. These data sets are regarded as uniformly distributed samples, since most cryptographically secure practical random number generators output almost uniformly distributed sequences. We use the NIST SP 800-90B entropy assessment python package of [4] to calculate the min-entropy of each data.

In Table 1, we can discover that the experimental minimum value, measured in bits, of optimum guessing attack cost and the assessed value of min-entropy for each data coincide with the result of NIST [3] and this paper.

5 Conclusion

The description of NIST [3] for the relationship between the min-entropy and the cost of guessing attack is a convincing argument to claim the legitimacy for the usage of min-entropy as a measure of cryptographic entropy sources. However the argument lacks the mathematical rigour, thus we provided an elaborate probabilistic analysis for this relationship. Our result contributes to the soundness of the relationship between the min-entropy and the guessing attack cost. We also provided some simulation results in order to investigate practicality of optimum guessing attack.

Acknowledgements. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. NRF-2014M3C4A7030648).

References

1. Cachin, C.: Entropy measures and unconditional security in cryptography. Ph.D. thesis, Reprint as vol. 1 of ETH Series in Information Security and Cryptography, ISBN 3-89649-185-7, ETH Zurich (1997)
2. Massey, J.L.: Guessing and entropy. In: Proceedings of IEEE International Symposium on Information Theory, p. 204 (1994)
3. NIST Draft Special Publication 800-90B (Second draft), Recommendation for the Entropy Sources Used for Random Bit Generation, 27 January 2016
4. https://github.com/usnistgov/SP800-90B_EntropyAssessment
5. http://ubld.it/products/truerng-hardware-random-number-generator/

Dynamic QoS Scheme for InfiniBand-Based Clusters

Bongjae Kim and Jeong-Dong Kim(✉)

Department of Computer Science and Engineering, Sun Moon University,
70, Sunmoon-ro 221 Beon-gil, Tangjeong-myeon, Asan-si
Chungcheonnam-do 31460, South Korea
{bjkim,kjd4u}@sunmoon.ac.kr

Abstract. Cluster-based computing systems are very widely used in various fields, including simulations and big data processing. InfiniBand is de-facto interconnect technology for cluster-based computing. QoS is a very important issue in data communication of cluster-based computing systems. In this paper, we propose dynamic QoS scheme for InfiniBand-based clusters. The proposed scheme can change QoS level in terms of the bandwidth. The proposed QoS scheme can get more bandwidth by changing the QoS level. The prototype of the proposed scheme was implemented in a real InfiniBand-based clusters. By using the prototype implemented, we confirmed and evaluated the usefulness and effectiveness of the proposed scheme.

Keywords: InfiniBand · Cluster · QoS · Bandwidth · High performance computing

1 Introduction

Recently, cluster-based computing systems are very widely used in various computing areas due to its massive computing power [1]. InfiniBand is a de-facto technology which is used to interconnect each node of cluster-based computing systems [2, 3]. In InfiniBand-based clusters, each computing or storage node is connected with InfiniBand Host Channel Adapter (HCA) and InfiniBand Switches. The data rate is different according to the technical specifications. For example, EDR (Enhanced Data Rate) technology supports data transfer rate of 100 Gb/s.

InfiniBand supports a memory-based user-level communication and RDMA (Remote Direct Memory Access) communication. In the InfiniBand, the communication interface is called *Queue Pair* (*QP*), which is the logical endpoint of a InfiniBand communication link. A QP is implemented on the host side of an InfiniBand HCA. Each InfiniBand HCA can have multiple ports. The port side of the InfiniBand HCA implements Virtual Lanes (VL). Virtual lanes support multiple independent data flows over physical communication link. InfiniBand typically supports between one and 15 virtual lanes. In addition, there are Service Levels in InfiniBand architecture. SL used to indicate a class of service [4]. Each InfiniBand switch in the cluster maintains a mapping table between the SL and VL. Based on the mapping table, InfiniBand switch selects a certain output port virtual lane. As we explained before, InfiniBand communication

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_88

links are logically split into multiple virtual lanes. Each VL has its dedicated set of resources for data buffering. We can assign a weight value to each VL. Based on the weighted value, weighted fair arbitration is used to transfer data [5]. In short, the weighted value specifies the amount of bandwidth allocated.

QoS (Quality of Service) is a very important and key issue in data communication of cluster-based computing systems. Generally, QoS means that the capability of a network to support better service to selected data traffic or applications over interconnect network technologies. The typical main goal of the QoS is to provide priority to selected data traffic in terms of bandwidth or latency. However, original InfiniBand only supports static QoS functionality.

In this paper, we propose dynamic QoS scheme over InfiniBand-based clusters. Our scheme can change its QoS level when transferring data. In shorts, our scheme changes its service level dynamically while considering the estimated current bandwidth and its minimum requirements. The proposed scheme was implemented in a real InfiniBand-based clusters. We evaluated the performance of proposed scheme. Based on the experimental results, the proposed scheme can provide dynamic QoS function when transferring the data.

Rest of this paper is organized as follows. In Sect. 2, we will explain the our dynamic QoS scheme in detail. In Sect. 3, we will show the performance evaluation. Finally, we will conclude this paper with future works in Sect. 4.

2 Dynamic QoS Scheme Over InfiniBand

Generally, InfiniBand's managed switches have a build-in subnet manager. In our dynamic QoS scheme, we used software-based OpenSM (Open SubnetManager) rather than build-in subnet manager to manage and maintain the cluster. To assign the weight value and make mapping table between the SL and VL, we have to modify the QoS policy file of OpenSM. When QoS functionality of OpenSM is enabled, OpenSM looks for a QoS policy file.

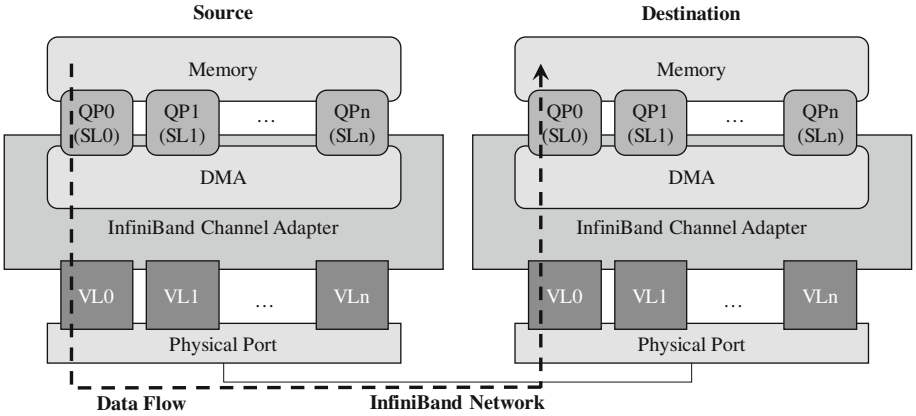
Table 1 shows the parameters and its descriptions for SL2VL mapping and VL arbitration [6]. In the QoS policy file, we should modify the *qos_vlarb_low* field. The below code shows an example of some of modified QoS policy file for the OpenSM. In this example, VL1 is been allocated 70 %, VL2 is 20 %, and VL3 is 10 % because VL1's weight is 224 and the summation of weights of each VL is 320. By using this policy, we can support QoS for each data traffic. However, original InfiniBand and OpenSM only support static QoS.

An example of some of modified QoS policy file for OpenSM

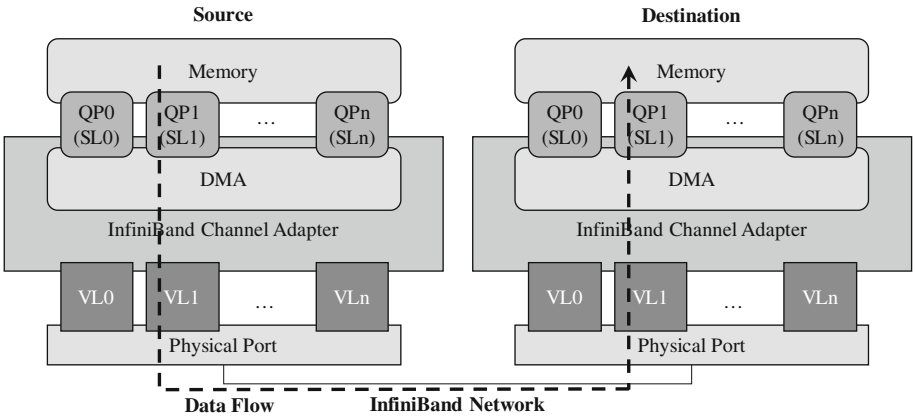
```
Qos_max_vls 8
qos_high_limit 0
qos_vlarb_high 0:32
qos_vlarb_low 1:224, 2:64, 3:32
qos_sl2vl 0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15
```

Table 1. Parameters and its descriptions for SL2VL mapping and VL arbitration.

Parameters	Descriptions
Max VLs	The maximum number of VLs that will be on the subnet
High limit	The limit of High Priority component of VL Arbitration table
VLArb low table	Low priority VL Arbitration table
VLArb high table	High priority VL Arbitration table
SL2VL	SL2VL Mapping table. It is a list of VLs corresponding to SLs 0-15



(a) Before changing QP



(a) After changing QP

Fig. 1. The concept of the proposed dynamic QoS scheme

In our dynamic QoS scheme, we make multiple QP for a logical data communication link between source and destination to support dynamic QoS. Each QP has different service level over VLs. If estimated bandwidth is lower than the requirement in terms of bandwidth, our proposed QoS scheme changes current QP to other QP,

which uses more high service level than current QP. Therefore, a certain traffic can get more bandwidth by changing its QP.

Figure 1 shows the concept of the proposed dynamic QoS scheme. In this Fig. 1, we assume that SL 1 has more weighed value than SL 0. Therefore, if we use SL 1 rather than SL 0, we can get more network bandwidth. As shown in Fig. 1, the proposed QoS scheme can change the QP used in the communication from QP 0 to QP 1 in order to get more bandwidth. It is also very important to assign a proper weight value to each SL and VL. We do not address this issue in this paper.

3 Performance Evaluations

In this section, we will explain about the performance evaluation of the proposed dynamic QoS scheme. First of all, we will explain the evaluation environment. At last, we will show the evaluation results of the prototype of the proposed dynamic QoS scheme.

3.1 Evaluation Environments

Table 2 shows the computing environments used in the performance evaluations. InfiniBand Host Channel Adapter (HCA) which supports 56 Gb/s is used to interconnect each computing node. Each computing node is equipped with two Intel Xeon E5-2650 CPUs and 64 GB memories. The version of OpenSM is 4.3.0. Operating system is Red Hat Enterprise Linux 6.5 and its version is 2.6.32.

Table 2. InfiniBand-based cluster computing environment.

Features	Descriptions
CPU	2 × Intel Xeon CPU E5-2650 2.6 GHz v2
Memory	64 GB
OS	Red Hat Enterprise Linux 6.5
Kernel version	2.6.32
OpenSM	4.3.0
OFED version	OFED-2.1-1.0.6
InfiniBand HCA	Mellanox ConnectX-3 VPI Adapter Dual-Port QSFP, 56 Gb/s
InfiniBand switch	Mellanox SX6018 18 Port 56 Gb/s InfiniBand/VPI Switch

Table 3. Weight values for each SL used in the experiment

Service level	Weight value	Allocated bandwidth
1	224	70 %
2	64	20 %
3	32	10 %

Table 3 shows the weight values used in the experiment. Service level 1, 2, and 3 are allocated 70 %, 20 %, and 10 % of total bandwidth, respectively. In other words, if there are three traffics which use SL1, SL2, and SL3, respectively, the traffic which uses SL1 will get 70 % of total bandwidth during communication.

3.2 Evaluation Results

In the experiment, we used two data transfer traffics at the same time. In other words, there are two network traffics over one physical InfiniBand link. First one only uses SL 1 when transfer data. Second one dynamically changes its service level in order to get more bandwidth while considering its estimated bandwidth and QoS requirement.

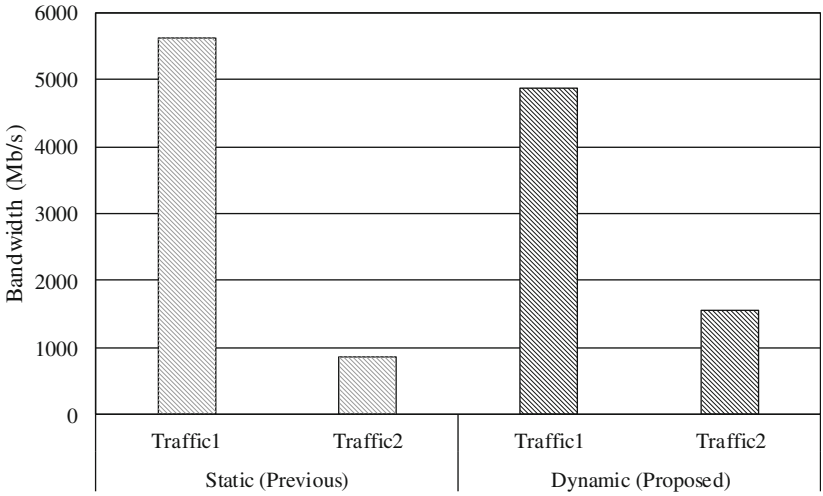


Fig. 2. Bandwidth comparison between static and dynamic QoS

Figure 2 shows the bandwidth results between previous static and proposed dynamic QoS. The traffic 2 of proposed QoS scheme can change its service level to acquire more bandwidth to meet QoS requirement. As shown in Fig. 2, the traffic 2 can get more bandwidth up to 1,560 Mb/s when compared to the static QoS scheme. On the other hand, the bandwidth of the traffic 1 was decreased to 4,880 Mb/s from 5,611 Mb/s. The summation of bandwidths of each traffic is about 6,449 Mb/s. Based on the experimental results, our proposed QoS scheme can support dynamic QoS in terms of bandwidth by changing service level.

4 Conclusion and Future Works

QoS is a very important issue in data communication. In this paper, we proposed dynamic QoS scheme over InfiniBand-based clusters. Our dynamic QoS scheme can change the QoS level in terms of data transfer bandwidth. The prototype of the

proposed scheme was implemented in the real InfiniBand-based cluster computing environment. Based on the performance evaluation results, we could confirm that the proposed scheme can service dynamic QoS over InfiniBand-based clusters.

In the future works, we will adjust the proposed QoS scheme to real cluster-based applications and evaluate the performance in terms of QoS and execution time.

Acknowledgments. This work was supported by the Sun Moon University Research Grant of 2016.

References

1. Strohmaier, E., Dongarra, J.J., Meuer, H.W., Simon, H.D.: Recent trends in the marketplace of high performance computing. Parallel Comput. **31**(3), 261–273 (2005)
2. InfiniBand Trade Association: InfiniBand Architecture Specification: Release 1.0. InfiniBand Trade Association (2000)
3. Low, J., Chrzeszczyk, J., Howard, A., Chrzeszczyk, A.: Performance assessment of infiniband HPC cloud instances on Intel Haswell and Intel Sandy bridge architectures. Supercomput. Front. Innov. **2**(3), 28–40 (2015)
4. Reinemo, S., Skeie, T., Sodring, T., Lysne, O., Torudbakken, O.: An overview of QoS capabilities in infiniband, advanced switching interconnect, and ethernet. IEEE Commun. Mag. **44**(7), 32–38 (2006)
5. Crupnicoff, D., Das, S., Zahavi, E.: Deploying quality of service and congestion control in infiniband-based data center networks, (White Paper). Mellanox (2005)
6. Mellanox: Mellanox OFED for Linux User's Manual. Mellanox (2011)

Applying PE-Miner Framework to Software Defined Network Quarantine

Dong-Hee Kim[1(✉)], Soo-Hwan Lee[1], Won-Sik Doo[1], Sang-Il Ahn[2], and Tai-Myoung Chung[3]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea
{kkim, drleesh, stuartdoo}@imtl.skku.ac.kr
[2] Department of Information Seucrity, Beaksoek University, Cheonan, Korea
asi9203@bu.ac.kr
[3] College of Software, Sungkyunkwan University, Suwon, Korea
tmchung@skku.edu

Abstract. As increasing the size of network, the malware propagates to other network easily. Moreover, malware is hard to detect if it is modified. The complexity of current network also causes the weakness for malware detection. Therefore, SDN quarantine network architecture has been researched. We applied the improved PE-miner framework that is malware detection mechanism based on machine learning algorithm to the SQN 1st quarantine. 1st quarantine is the system that filtering the malware using static mechanism. In this paper, detection rate of the improved PE-miner framework was evaluated and the real-time performance was also tested. Referring the result, we have proved that applying the PE-miner framework to SQN 1st quarantine is permissible.

Keywords: Software defined network · Malware detection · Machine learning

1 Introduction

Software Defined Network (SDN) is an emerging network architecture that the physical separation of the network control plane from the forwarding plane, and where a control plane controls several devices [8]. SDN developed to solve three main problems in current networks [1]. First, support the automated management and centralization. When network environment had changes, for example if new firewall is installed, switches condition are changed, or new equipment deployed by network manager, the entire network policies or configurations should be changed. In SDN architecture, administrator focuses on managing the control plane. Second, improve the efficiency on Spanning Tree Protocol (STP) and Routing Protocol. SDN control plane know the entire network resources conditions because of the centralized architecture. Therefore, controller decides the efficient routing path with consuming the low cost. Third, solve the synchronization problem. Current network architecture each network device handles the packet and it causes the synchronize overhead of each equipment. But, in SDN architecture, one superior controller handles the routing path. The Dijkstra algorithm,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_89

which is one of the routing protocols, is easy to implement in a centralized network comparing to current distributed network [1]. Furthermore, the SDN Quarantine Network (SQN) which is network security function collaborated with SDN network architecture has been suggested by Kim et al. [2].

Most of the security equipment, e.g. IDS, IPS, firewall, etc., are detecting the malware using the misuse detection mechanisms. Misuse detection mechanisms save the signature of malicious code or actions. Thus, it cannot detect the new malware or modified malicious code. Shafiq et al. [4]. have implemented malware detection framework using Portable Executable header (PE-header). The PE-miner framework uses the machine learning algorithms to detect malware. Portable Executable header (PE-header) is used to predict the executable files. This framework unnecessary to save all the signature of malware but it demands to study some samples. By training samples, it can detect not only recognized malware but also unknown one. Therefore, PE-miner framework is suitable to be mounted in SQN. But, we changed the number sample that trained, and used the Classification and Regression Tree (CART) instead of J48. CART is similar to C4.5 but it also supports numerical target value. Moreover, two results are evaluated; the prediction accuracy of PE-framework itself, and the prediction time depending on size of files in SQN environment.

2 SDN Quarantine Network (SQN)

SQN is a new quarantine method using the SDN environment. Figure 1 represents the SQN architecture. SQN composed with SQN Controller, SQN Switches, 1st Quarantine, 2nd Quarantine, Clean Area Manager, and Agents. SQN Controller represents the SDN controller which has responsibility for network topology management, routing path calculation, and deciding routing QoS [3]. 1st and 2nd quarantines have a role to detect the malware using static and dynamic methods. Clean Area Manager collects the reports from Agents and then analyzes new pattern of malware. Agents monitor the suspicious process and report these to manager if malicious act has been initiated.

SQN has three areas. One is Internet area, the place where potential threats are existing. Viruses, worm, and Trojans are coming from unsecured source. Second, quarantine area is the place where checking data that came from internet. In this area, split packets which are forwarded by SQN Controller are reassembled by 1st quarantine. These data are checked whether malicious or benign through static way. Detecting methodology is proposed in previous research [5]. They link the various malware detecting modules including machine learning mechanisms and each module calculates the score that represents probability of malicious code. If the score is low enough, data is forwarded to the agent. If score value that is over the threshold, the corresponding data is discarded. If data is judged as a suspicious one, in case that score does not meet the minimum value at the same time it is not exceed threshold, it is sent to the 2nd quarantine. The 2nd quarantine analyzes the corresponding data in dynamic way. If 2nd quarantine disbelieves the data again, suspicious tag is attached to the packets and is sent to agents. For the last, clean area is local network. Agents are the special clients that have a responsibility to monitoring the suspicious data. The suspicious data is observed the system call and file access record by Agent. If suspicious

data was malware, the agents report the log to the Clean Area Manager. Clean Area Manager analyzes the log and informs the quarantine systems. Quarantine systems use the report to update their malware information. Therefore, this report is the requisite element for keeping the stability of entire network. The PE-miner framework is applied as one of the 1st quarantine module and it evaluates the executable data.

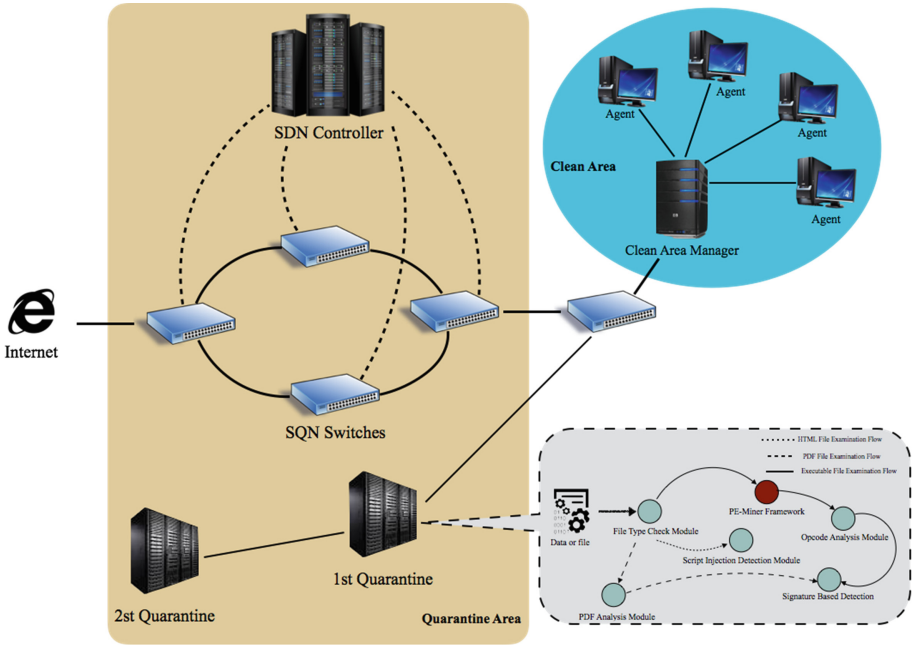


Fig. 1. SQN architecture

3 PE-Miner Framework

PE-miner framework is developed by Shafiq et al. [4]. PE-miner framework has a detection mechanism for malicious executable using PE-header. Three preprocess mechanisms are used before adapting to machine learning algorithms.; Redundant Feature Removal (RFR), Principal Component Analysis (PCA), and Haar Wavelet Transform (HWT). Among these, the HWT has derived best performance for detecting the malware. HWT stores the most relevant information with the highest coefficients at each order of a transform. The lower order coefficients have possible to be ignored for get only the most relevant information [4]. And the various machine learning algorithms are used. They use J48 (decision tree), IBk (in- stance based learner), NB (Native Bayes), RIPPER (inductive rule learner), and SMO (support vector machines using sequential minimal optimization) for malicious executable detection. The J48 outperformed with more than 99 % of detection rate and less than 0.5 % false-positive rate when 11,786 training sample used. Thus, we have decided to using a decision tree

algorithm for 1st quarantine module. However, we applied the CART instead of J48. CART is similar to C4.5, but threshold that yields the largest information gain at each node [7].

3.1 Learning Data

We have collected the 5,000 benign samples and 10,000 malicious samples to compose the learning data. Benign samples were collected from the 'system32' folder of Windows 7. Malicious codes were downloaded from the 'VXheaven.org' [6]. The learning data is a key component of machine learning mechanism. For generating the learning data, 15,000 samples were used. Furthermore, for improving the accuracy, the characteristics should be well-defined. We referred the characteristics of PE-header that previous researches defined. But we only use some attributes, these are Characteristics, Number of Symbols, Major Linker Version and Size of Initial Data, Major Image Version, and the DLL Characteristics. According to our evaluation, other attributes value that evaluated from previous research is now featureless.

4 Test Result and Evaluation

We have experiment module for two tests. One is to know the efficiency of module itself. Another is verifying the real-time guarantee of the module adapted in SQN. For the module test, learning data is used trained to CART algorithm and the 50,000 malicious data is passed to check the error rate. For the real-time experiments, we randomly choose the 10 sample files depending on the size of the file. We have checked the average time when 10 files are received. The module experiment is done with Core i5 3.30 GHz processor with 16 GB RAM. Quarantine is Intel Xeon 2.60 GHz processor with 64 GB RAM.

The use of CART algorithm performance result is in Table 1. The PE-miner properly classifies all the samples that were used for making the learning data. Also for unrecognized data, the detection accuracy represents more than 99 % with 0.08 % of false-positive rate. This performance result is advanced than the J48 algorithm (J48 algorithm shows 0.5 % of false-positive rate). However, the benign sample detection rate is improperly performed. It shows 96 % accuracy with 3.85 % error. The accuracy is measured as low, but in SQN it is acceptable. Because not only this framework decides the malware packets but also other modules do. The rating ratio is decided by administrator of SQN. The result is recommended to administrator that giving a higher score when detecting malware.

On the other hands, the average of required time for detection is shown in Fig. 2. Time is checked when all of the 10 files are delivered. As an evaluation result represents, required time increase as long as size of file increasing. Especially in range of file size 501–600 and 901–1000 shows notable growth value. Even though required time is increasing by size of files, but still the time value shows around 0.1 s. And also it is the required evaluation time for checking the all 10 files. Detecting one file should not require much time. Thus, the evaluation time is reasonable for the real-time efficiency in SQN.

Table 1. PE-miner with CART performance result

Sample used for training	Unrecognized data	
	False-positive	False-negative
100 % (All Classified)	0.08 %	3.85 %

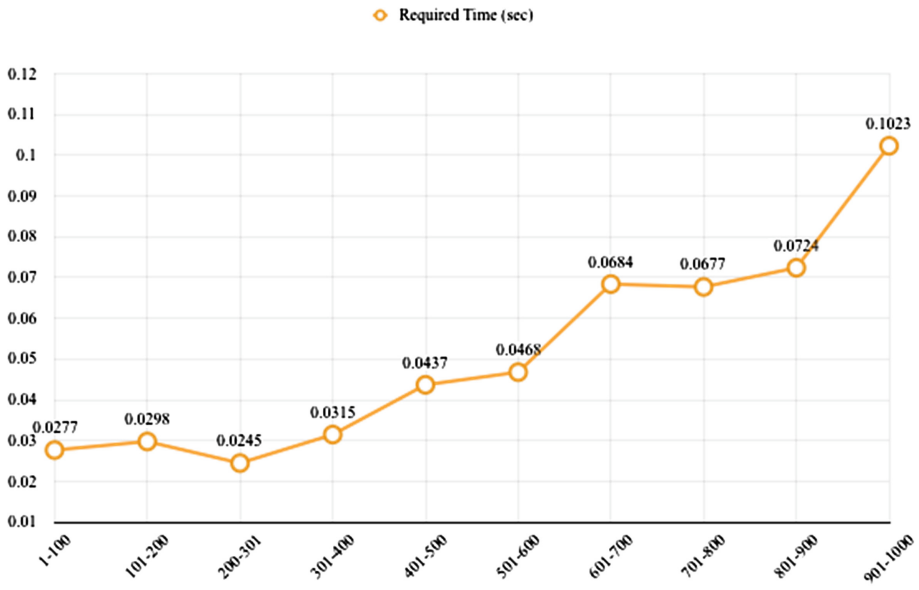


Fig. 2. Required time for detecting 10 files

5 Conclusion

We have adapted the PE-miner framework to SQN 1st quarantine. Only some reasonable attribute is selected for making training data. Also, CART algorithm is used rather than J48. Training data is composed with 10,000 malicious executables and 5,000 benign executables. The experimental result is obtained with 50,000 malware samples. The result denotes that 99 % accuracy with 0.08 % of false-positive error for detecting malware. It was improved approximately 6 times comparing to the PE-miner framework. Moreover, PE-miner framework is affected by types of malware. According to the previous research, the filtering rate seems low when classifying the Trojans and Worms. But, improved PE-miner framework that is our model was not concerned the types of malware. In addition, filtering accuracy is reasonable enough in SQN 1st quarantine mechanism. Because not only PE-module filters the malware but also other modules are detecting the malware. The filtering time is increasing as long as file size increase. But, the required time is derived 0.1 s with 1000 KB files. In addition, the test was evaluated with the 10 files. It means that the required time for filtering is more less than 0.1 s. Therefore, the filtering time does not affect the real-time efficiency.

Acknowledgments. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R-20160222-002755, Cloud based Security Intelligence Technology Development for the Customized Security Service Provisioning). And also this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0020210).

References

1. Seo, Y., Lee, M.: Introduction to SDN. Youngjin, Seoul (2014)
2. Kim, N., Jung, J., Song, Y., Kim, H., Jung, T.: The Design of SDN Quarantined Network. Electron. Inf. Res. Inf. Center. 559–560 (2014)
3. Nadeau, T., Gray, K.: SDN
4. Shafiq, M., Tabish, S., Mirza, F., Farooq, M.: PE-miner: mining structural infromation to detect malicious executables in realtime. In: International Workshop on Recent Advances in Intrusion Detection, pp. 121–141 (2009)
5. Kim, D., Jung, J., Chung, T.: The architecture of detecting malicious behavior in SQN quarantine using static analysis. In: Korea Society of Digital Industry and Information Management, pp. 102–105 (2015)
6. Virus collection (VX heaven). http://vxheaven.org/vl.php
7. Decision Trees documentation. http://scikit-learn.org/stable/modules/tree.html#tree-algorithms
8. Sezer, S., Hayward, S., Chounhan, P.: Are we ready for SDN? Implementation challenges for software-defined networks. IEEE Commun. Mag. **51**, 36–43 (2013)

A Novel Method for Eliminating Duplicated Frames in Ethernet Standard (IEEE 802.3) Networks

Saad Allawi Nsaif and Jong Myung Rhee(✉)

Department of Information and Communications Engineering,
Myongji University, Yongin, South Korea
saad.allawil@gmail.com, jmr77@mju.ac.kr

Abstract. If assumed that each Ethernet standard (IEEE 802.3) node has more than one port connected to the network and the node duplicates each sent frame, then an active or a seamless redundancy will be established because the destination node will receive at least two frame copies with zero recovery time. In this paper, we present a novel method for eliminating the duplicated frames in any Ethernet network type. This will ensure that the destination node will only consume one frame copy from each sent frame and eliminate the other copies. The proposed method will set a counter on each receiving port in the destination node. The destination node will consume the copy that arrives through the fastest path, or in other words, through the port that has the fastest counter value. The proposed method does not need to disable ports as required by rapid spanning tree protocol (RSTP) or the media redundancy protocol (MRP), to avoid looping issues; instead, it activates all ports to provide a type of better traffic distribution among the network links.

Keywords: Elimination of duplicated frames · Discarding redundant frames · Seamless redundancy · Active redundancy

1 Introduction

In IT and communication sectors, fault tolerance plays an important role, especially the type that provides seamless redundancy, which is required in mission-critical applications, such as future smart cars (driverless cars). This type of car requires seamless redundancy in the control system to avoid any interruptions in control signals, such as brake signals, because a serious accident could occur otherwise. Other industrial automation applications would also need to adopt the seamless redundancy concept in control systems, such as substation automation system (SAS) networks. Since the Ethernet standard (IEEE 802.3) [1] does not provide a fault-tolerance capability, various protocols for industrial automation systems have been developed and standardized by the International Electrotechnical Commission (IEC) and the Institute of Electrical and Electronics Engineers (IEEE) [2], such as rapid spanning tree protocol (RSTP) as IEEE 802.1D [3], media redundancy protocol (MRP) as IEC 62439-2 [4], cross-network protocol (CRP) as IEC 62439-4 [5], beacon redundancy protocol (BRP) as IEC 62439-5 [6], parallel redundancy protocol (PRP) as IEC 62439-3-clause 4 [7], high-availability

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_90

seamless redundancy (HSR) protocol as IEC 62439-3- clause 5 [7]. Among these protocols, only two protocols, PRP and HSR protocols, are suitable for seamless communication. Both HSR and PRP are based on the same principle of active redundancy by duplicating the information frames, which therefore results in a zero-delay reconfiguration in the case of a switch or link failure [7, 8]. The RSTP that is used with the Ethernet standard to provide path redundancy usually requires an upper bound of 2 s per switch [8] because once a link or a switch fails, the network undergoes reconfiguration to rebuild the logical paths using the RSTP. This makes the RSTP unsuitable for industrial applications that require zero-recovery time or close to that, such as IEC 61850 standard's messages that require 3–4 ms as a maximum timeout. The HSR protocol provides two frame copies for the destination node from each sent frame, one from each side, enabling zero-fault recovery time if one of the frame copies is lost. This means that even in the case of a node failure or a link failure, network operations are not interrupted. If both sent copies reach the destination node, the node will consume the fastest copy and discard the other copy. This feature of the HSR protocol makes it very useful for time-critical and mission-critical systems; however, the HSR and the PRP protocols are required to add a header of 6 bytes in length in the Ethernet standard frame to distinguish between redundant frame copies of the same sent frame. This addition makes the new Ethernet frame unsuitable for Ethernet standard off-shelf end user devices, such as PCs or printers.

In this paper, we introduced a novel method to eliminate the duplicated frame copies in any Ethernet standard network to ensure that the destination node will consume one frame copy from each sent frame and delete other copies that might be delivered later. This will be useful in any Ethernet network when each of its nodes duplicates each sent frame. By this way, an active or a seamless redundancy with zero recovery time will be provided. In other words, instead of disabling ports; as it is in RSTP or MRP to prevent looping issues; in our idea each sending node will duplicate each sent frame, consequently, the destination node will receive more than one copy from each sent frame; consume the fastest and delete the others. In case of one of these sent frames is lost, the destination node will not face any interruption or frame lost as it was in RSTP or MRP. In our method, the Ethernet network also does not need to disable ports to prevent looping. On the contrary, all ports will be active. The basic idea of the proposed method was introduced in [9], but for this paper, the elimination process for the duplicated frames is described in details.

The remainder of the paper is organized as follows. In Sect. 2, the process of eliminating duplicated frame is described in details under failure and failure-free cases. In Sect. 3, conclusions and suggestions for future work are presented.

2 Process of Eliminating Duplicated Frames

At the beginning, in order to enable the Ethernet standard (IEEE 802.3) protocol to provide active or seamless redundancy, it is required to assume that each Ethernet node shall has more than one port connected to the network and send duplicated frames for each sent frame. The destination node will receive duplicated frame copies from each sent frame, therefore, the node will need a method to consume one of each duplicated

frame copies and to eliminate the others. The standard HSR protocol is based on the same concept; however, it involves comparing the source MAC address and the sequence number in the frame's header of each copy to recognize whether the received frame has been received before. However, because the frame format of the Ethernet standard does not have headers inserted in the frame, it needs to follow an alternative method to recognize the duplicated frame copies. The proposed method of this paper will solve this issue as illustrated in the network example shown in Fig. 1 for both failure-free and failure cases.

2.1 Failure-Free Case Under Unicast Traffic

Assume node A sends many frames to node D. Node A will duplicate each sent frame and send them out, one copy into a direction. The Ethernet protocol is a time synchronized approach, which means that at each clock time, the frame copies will move one hop. Therefore, the following steps will occur during each clock time.

(a) First Clock Time

Node B will receive the first frame copy, and then it will do the following:

- Read the destination and the source MAC addresses, which are nodes D and A respectively in this case, and then establish a counter on each of node B's ports: port-a and port-b.
- These counters will only be associated with the connection pair (source-destination) of the source node A and the destination node D. In other words, these counters will only be increased if any has received a frame that originated from node A with the destination node D. In this case, the counter of port-a will be set as follows:

$$\alpha(B)\ a_A^D = 1,\ \alpha(B)\ b_A^D = 0,\ \textit{Clock time} = 1$$

where $\alpha(B)\ a_A^D$ is the counter of node B on port-a that is associated with the source node A and the destination node D.

- Since node B is not the destination of the frame, in the next clock time, node B will forward the received frame to the next node.

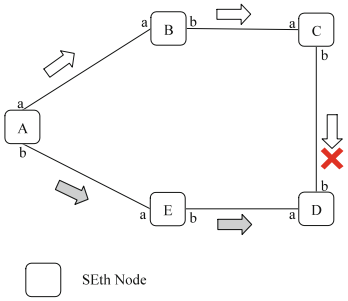


Fig. 1. Proposed method's behavior under a unicast traffic type.

During this clock time, clock no. 1, node E will complete the same steps as node B regarding the received frame copy.

(b) Second Clock Time

During the second clock time, or the first clock time for node D (from the perspective of port-a), node D will receive a frame copy through port-a and then perform the following:

- Read the destination and the source MAC addresses of the received copy and establish a counter on port-a and port-b. These counters will be associated with the connection pair node A- node D.
- Increase the counter of port-a by one, whereas port-b's counter will remain at zero as follows:

$$\alpha(D)\ a_A^D = 1,\ \alpha(D)\ b_A^D = 0,\ \textit{Clock time} = 2$$

- Node D will determine that it is the destination for that frame; therefore, it will consume it by sending it to the upper layer.

During this clock time, node C will also receive a frame copy through port-a. Node C will complete the same steps as node D, but instead of consuming the frame copy, it will forward the received frame copy to node D through port-b during the next clock time.

$$\alpha(C)\ a_A^D = 1,\ \alpha(C)\ b_A^D = 0,\ \textit{Clock time} = 2$$

(c) Third Clock Time

Node D will receive the second copy that was forwarded by node C. This copy will be received through port-b. However, node D will complete the following steps:

- Read the destination and the source MAC addresses, which are node D and node A respectively and then it will only update the values of its counters according to the information of the received frame.
- Since the frame copy is received from port-b, node D will increment the counter of port-b by one. Now, both of the counters are equal, which means they received the same number of frames—one frame from each port. In other words, the node received a redundant frame and will delete it.

$$\alpha(D)\ a_A^D = 1,\ \alpha(D)\ b_A^D = 1,\ \textit{Clock time} = 3$$

Since $\alpha(D)\ a_A^D$ was one and then $\alpha(D)\ b_A^D$ was incremented to one, $\alpha(D)\ b_A^D$ has received a redundant frame because $\alpha(D)\ a_A^D$ was greater than $\alpha(D)\ b_A^D$ before receiving the second copy. In other words, port-a is faster than port-b. Consequently, node D will always consume the first copy from port-a and delete the second copy delivered

through port-b. Therefore, node D will delete the second copy delivered through port-b and will keep using the same way with the other remaining sent frames.

2.2 Failure Case Under Unicast Traffic

For the same network shown in Fig. 1, assume node A sends four frames to node D.

(d) First Clock Time

Node E will receive the first copy of the first frame through port-a. Node E will forward it to node D after it finds that this frame's destination is node D; however, node E's counters will be:

$$\alpha(E)\ a_A^D = 1,\ \alpha(E)\ b_A^D = 0,\ \textit{Clock time} = 1$$

Node B will receive the second frame copy of the first frame and complete the same steps as node E.

$$\alpha(B)\ a_A^D = 1,\ \alpha(B)\ b_A^D = 0,\ \textit{Clock time} = 1$$

(e) Second Clock Time

Node E will receive the first copy of the second frame, and then it will increment port-a's counter. Assume the frame was corrupted by a bad cyclic redundancy check (CRC) value. Node E will delete and remove it from the network.

$$\alpha(E)\ a_A^D = 2,\ \alpha(E)\ b_A^D = 0,\ \textit{Clock time} = 2$$

Node B will receive the second copy, update port-a's counter value, and then forward the received frame to the next node during the next clock time.

$$\alpha(B)\ a_A^D = 2,\ \alpha(B)\ b_A^D = 0,\ \textit{Clock time} = 2$$

Node D will receive the first copy of the first frame. The node will complete the following steps:

- Read the destination and the source MAC addresses and then establish counters.
- Increment port-a's counter and consume the first copy by sending it to the upper layers.

$$\alpha(D)\ a_A^D = 1,\ \alpha(D)\ b_A^D = 0,\ \textit{Clock time} = 2$$

Node C will receive the second copy of the first frame and will complete the same steps as node D, but instead of consuming the copy, it will forward it to node D.

$$\alpha(C)\ a_A^D = 1,\ \alpha(C)\ b_A^D = 0,\ \textit{Clock time} = 2$$

(f) Third Clock Time

Node E will receive the first copy of the third frame, increment port-a's counter, and then forward that frame to node D.

$$\alpha(E)\ a_A^D = 3,\ \alpha(E)\ b_A^D = 0,\ \textit{Clock time} = 3$$

Node B will complete the same steps with the second copy of the third frame.

$$\alpha(B)\ a_A^D = 3,\ \alpha(B)\ b_A^D = 0,\ \textit{Clock time} = 3$$

During this clock period, which is the second from the node D's port-a perspective, node D should receive the first copy of the second frame through port-a, but since node E has deleted it because it was corrupted, node D will not receive any frame during this clock period. Therefore, it will determine that there is a missed frame in this clock and will wait for the second copy of the second frame that should be delivered through port-b. Consequently, node D will build or update a table called a lost frames (LF) table. The LF table lists all lost frames that are associated with each source node, clock time, and the port ID, as shown in Table 1.

Therefore, node D will consume the copy of port-b during the second clock time from port-b's perspective. Node D will also receive the second copy of the first frame through port-b, increment port-b's counter, and then delete it because both counters have the same value, which means that the copy is redundant and has been received before.

$$\alpha(D)\ a_A^D = 1,\ \alpha(D)\ b_A^D = 1,\ \textit{Clock time} = 3$$

Node C will receive the second copy of the second frame, increment port-a's counter, and then forward it to node D.

$$\alpha(C)\ a_A^D = 2,\ \alpha(C)\ b_A^D = 0,\ \textit{Clock time} = 3$$

(g) Fourth Clock Time

Node E will receive the first copy of the fourth frame, increment port-a's counter, and then forward it to node D.

$$\alpha(E)\ a_A^D = 4,\ \alpha(E)\ b_A^D = 0,\ \textit{Clock time} = 4$$

Node D will receive the first copy of the third frame, increment port-a's counter, and then consume it. Node D will also receive the second copy of the second frame through port-b.

Since this is the second clock time from the perspective of port-b, node D will determine that this copy should be consumed because it is listed in its LF table. Therefore, node D will consume it and increment port-b's counter; however, node D will continue to receive other frame copies through port-a and delete the other copies of port-b because:

$$\alpha(D)\ a_A^D = 3 > \alpha(D)\ b_A^D = 2,\ \ Clock\ time = 4$$

Later, node D will clear the LF table after receiving the lost frame through port-b.

Table 1. LF Table of node D.

Source node	Clock time number	Port
Node A	2	a

2.3 Failure-Free Case Under Multi/Broadcast Traffic

For the same network shown in Fig. 1, if it is assumed that node A has broadcasted a single frame, the first copy will travel through port-a, and the second copy will travel through port-b. Nodes B and E will receive these copies and do the following:

- Receive the first and the second copy of the sent frame.
- Read the destination and the source MAC addresses.
- Establish counters on both ports and increment port-a's counter.
- Make a copy from the received frame.
- Forward the original copy to the opposite port, port-b.
- Calculate the CRC code to ensure that the received copy is error-free, and then send it to the upper layers, else delete it and wait to receive the other copy from the opposite direction.

Nodes C and D will complete the same steps as nodes B and E; however, when nodes C and D forward the first and second copies to the opposite port—port-b— node D will receive the second copy through port-b, and it will notice that the copy was sent from node A. Therefore, it will:

- Increment the counter of port-b.
- Compare the counter values of ports a and b that are associated with node A. Node D will notice that both values are equal, which means port-a has received this frame before. Therefore,
- Node D will delete the second copy because it is a redundant frame copy.

Node C will complete the same steps as node D regarding the copy delivered through port-b.

2.4 Failure Case Under Multi/Broadcast Traffic

In this case, the Ethernet standard will have the same behavior when encountering a failure case under a unicast traffic type.

3 Conclusion

The proposed method does not need to disable ports, as required for RSTP or other protocols, to avoid looping issues. The method relies on the ports' counters to delete the duplicate frame copies and all ports will be active, which provides a type of traffic balancing that is better than the Ethernet standard with RSTP protocol because the nodes will use them to send and consume traffic from a direction once and from the other direction once again. The method uses the same Ethernet standard frame layout, which is another advantage from economical and manufacturing perspectives. The method provides very high availability due to its seamless service, especially for time-critical applications, such as a substation automation system (SAS), smart cars, and several industrial automation systems. In summary, this method will make the Ethernet standard provides active or seamless redundancy.

In future work, a type of "machine learning" algorithm should be added to provide the Ethernet network switches with "intelligence" regarding forwarding network traffic into different paths when they detect a bottleneck at a certain area of the network. On the other hand, the proposed method should run a counter for each port per sending node; however, the number of counters on each port will depend on the number of sending nodes, which requires particular temporary memory resources. To accomplish this, a comprehensive study is required to allocate the optimum size of memory to each switch. Another enhancement would be reducing the number of these counters and establishing one counter on each port per each receiving node instead of per sending node, which would reduce the running counters, further improve the node performance, and reduce the size of the required memory.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) and funded by the Ministry of Education (No. NRF-2015R1D1A1A02059506).

References

1. IEEE 802.3 standard: IEEE Standards for Ethernet (2012)
2. Rhee, J.M., Nsaif, S.A.: Traffic Reduction Algorithms for High-Available Seamless Redundancy (HSR) Protocol, 1st edn. Press of Myongji Research Institute for DU, Yongin (2013)
3. IEEE 802.1D standard: IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges (2004)
4. IEC 62439-2 standard: Industrial Communication Networks: High-Availability Automation Networks, Part 2: Media Redundancy Protocol (MRP) (2016)

5. IEC 62439-4 standard: Industrial Communication Networks: High-Availability Automation Networks, Part 4: Cross-Network Redundancy Protocol (CRP) (2010)
6. IEC 62439-5 standard: Industrial Communication Networks: High-Availability Automation Networks, Part 5: Beacon Redundancy Protocol (BRP) (2016)
7. IEC 62439-3 standard: Industrial Communication Networks—High-Availability Automation Networks, Part 3: Parallel Redundancy Protocol (PRP) and High-Availability Seamless Redundancy (HSR) (2016)
8. IEC Technical Report: Network Engineering Guideline for Communication Networks and Systems in Substations, IEC 61850-90-4 (2013)
9. Nsaif, S.A., Rhee, J.M.: Seamless ethernet approach. In: 2016 IEEE International Conference on Consumer Electronics (ICCE), pp. 385–388 (2016)
10. IEEE 1588 standard: IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems (2008)
11. Nsaif, S.A., Rhee, J.M.: Improvement of high-availability seamless redundancy (HSR) traffic performance for smart grid communications. J. Commun. Netw. **14**, 653–661 (2012)

A Study of Malicious Code Classification System Using MinHash in Network Quarantine Using SDN

Soo-Hwan Lee(✉), Myeong-Uk Song, Jun-Kwon Jung, and Tai-Myoung Chung

College of Information and Communication Engineering, SungKyunKwan University, Suwon, Korea
{drleesh, mwsong, jkjung}@imtl.skku.ac.kr, tmchung@skku.edu

Abstract. Thanks to the development of IT technology, information systems have been growing continuously. However, there are threats behind the convenience. There is a possibility of malicious users to steal sensitive information and malware can lead to social chaos by paralyzing the information systems. Several solutions to prevent these attacks have been introduced. In this paper, we introduce malware detection technique using Minhash and evaluate the performance of it and suggest the cyber quarantine system applied this technique. It contributes to detect not only known malware but unknown malware.

Keywords: Software Defined Network · MinHash analysis · Static analysis

1 Introduction

As increasing embedded systems and IoT devices, many users have a convenience from a variety of fields. However, this development of IT technology leads to some problems like leakage of personal information, paralyzed service system through the attacks. Especially leakage of personal information can result in secondary damage and service system paralysis may lead to social turmoil, such as banking disruptions, traffic network paralysis. There are several ways to attack systems and insert a malicious code into transmitted file. Because of the fast growth rate of malicious code, it is important to detect and defense malicious code.

To prevent attacks using malicious code, security systems use static analysis and dynamic analysis. Static analysis is a technique for analyzing the code of the program which does not operate. Static analysis is composed of anomaly detection which monitors the pattern of unusual data and misuse detection which identifies attacks by monitoring the pattern of the data that is estimated to be malicious. Static analysis is difficult to detect an unknown attack since it is a signature-based analysis. On the other hand, dynamic analysis is a technique for monitoring the program that is running to detect malicious behavior [1].

Minhash analysis is an analysis technique in the static analysis categories. Therefore, Minhash analysis can't detect an unknown attack easily. This paper proposes the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_91

way to detect an unknown attack by using Minhash analysis and manager of cyber quarantine system using Software Defined Network (SDN).

This paper is organized as follows. Section 2 describes the malware detection techniques, Minhash, cyber quarantine system, and explains the architecture of malicious code classification system using Minhash and implements and verifies the system in Sect. 3. In this paper, we propose to add the Minhash analysis on cyber quarantine system in Sect. 4. Finally, Sect. 5 concludes and outlines future work.

2 Related Work

2.1 MinHash

Minhash function is different from the conventional hash function. Conventional hash function makes a completely different hash values when 1 bit of original input data is changed. On the other hand, Minhash function leads to a similar output when original input data is changed little. Minhash function is a Locality Sensitive Hashing (LSH) based on the Jaccard Similarity Clustering [2].

Jaccard Similarity is represented by $\frac{|A \cap B|}{|A \cup B|}$ that means the similarity of the two sets A, B. Thus Jaccard coefficient value is between 0 and 1. LSH has a technique called Fuzzy Hashing. This technique divides the file into a fixed chunk size and obtains a hash value of each chunk. After this stage, it merges these hash value into one string. As a result, the hash value of the original data's modified part is changed only. Jaccard Similarity can measure the degree of similarity between two files using this principle [3].

2.2 SDN Quarantined Network (SQN)

Figure 1 shows architecture of cyber quarantine system. The cyber quarantine system consists of a SQN switch, a quarantine station, a clean area (local network). The quarantine has a preprocessor, a primary quarantine station, and a secondary quarantine station. The clean area has agent and manager.

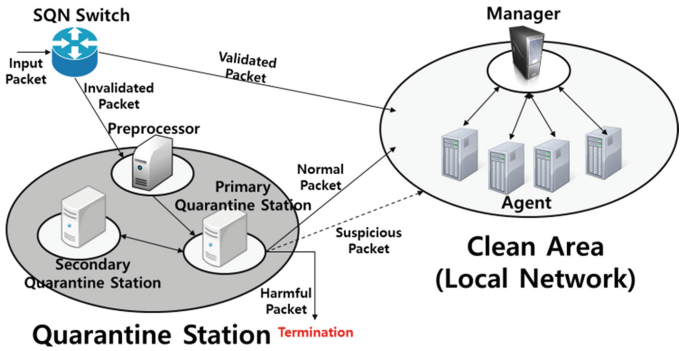


Fig. 1. Architecture of cyber quarantine system [4]

When the packets enter the cyber quarantine system through the SQN switch, the packets that are validated to normal file via packet authentication process at SQN switch are sent to clean area. Invalid packets are sent to quarantine station and reassembled at preprocessor. Reassembled packets are copied and transmitted to primary quarantine station. Primary quarantine station proceeds static analysis that classifies input packets as harmful, suspicious, normal packet. Normal packets are sent to clean area. In the case of suspected packets, primary quarantine station sends them to clean area after attaching the suspicious tag. Harmful packets are terminated immediately. If primary quarantine station can't classify the packets, they are transmitted to secondary quarantine station. Second quarantine station classifies the packets as harmful, suspicious, normal packet. Normal packets, suspicious packets are sent to clean area and harmful packets are discarded.

Manager and Agent in clean area share security policy. Agent monitors processes related the suspicious packets. When the processes behave an act departed from security policy, agent sends the event to manager. Manager can detect an unknown attack and establish a new policy by analyzing the event [5].

3 Malicious Code Classification System Using MinHash

3.1 Architecture of Malicious Code Classification System Using MinHash

Malicious code classification system using Minhash has the architecture such as Fig. 2.

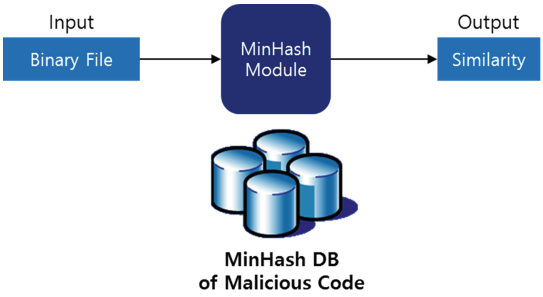


Fig. 2. Architecture of malicious code classification system using Minhash

In order to operate the malicious code classification system using Minhash, Minhash DB which is based on the existing malicious code must be created first.

When the files come in Minhash module, it extracts Minhash value. The module compares extracted Minhash value with all values in Minhash DB and measures similarity using Jaccard Similarity. When the similarity measurement is completed, the highest score becomes output value which shows the degree of similarity of input files with the existing malware.

3.2 Investigation of Malicious Code Classification System Using MinHash

Minhash function is Locality Sensitive Hashing (LSH) based Jaccard similarity. There is ssdeep tool based Fuzzy Hashing in LSH. In this paper, 8,280 of malicious code are entered into the malicious Minhash DB and compared the malicious codes with the Minhash DB. This experiment showed that it is possible to detect all of the malicious code that exists in Minhash DB of malicious code, because degree of similarity of all input malicious code is 100 %.

To check the false positives of Minhash function, 10,557 of benign code are entered into ssdeep tool and compared the code with Minhash DB. As a result, only one code has a degree of similarity that is determined to be 41 %.

We assume that because Minhash function is the function that output degree of similarity, it can detect unknown malicious code by measuring similarity of malicious code. Then we compare 100 of malicious code that does not exist in the Minhash DB of malicious code with the data stored in the DB. The results of this test are shown in the Table 1. In the table, 84 % of malicious codes unregistered in the Minhash DB have degree of similarity that is over 50 %. Because malicious code classification system using Minhash is a signature based static analysis, it has a limitation that can't detect all unknown malicious code.

Table 1. Degree of similarity and the number of input malicious code

Degree of similarity	The number of input malicious code
90–100	47
80–89	13
70–79	6
60–69	9
50–59	9
Less than 50	16

Through three experiments that compare malware registered in the DB, benign code, unknown malware with Minhash DB of malicious code, we can know the fact that Minhash function can detect malware registered in the DB. Assuming that the malicious code has the similarity which is over 50 %, it can detect unknown malicious code. Also, it has low false positive and false negative. Test results including false positive of can be found in the Table 2.

Table 2. Test result

		Actual status	
		Malicious code	Benign code
Test result	Decided to malicious code	8,280	1
	Decided to benign code	0	10,556

4 Cyber Quarantine System Applied MinHash Node

Malicious code classification system using Minhash can be applied to a cyber quarantine system. In the primary quarantine station, several static analyses will be conducted. It has graph form as shown in Fig. 3.

For example, 'PE header check', 'Header analysis', 'Opcode analysis' is quarantine module and a node in the primary quarantine station. Each node conducts n-gram analysis that analyzes the byte sequence of the file. It analyzes the appearance frequency of the opcode that is not frequently used [6]. There are nodes that execute static analysis using information such as file size, magic number and file creation time [3].

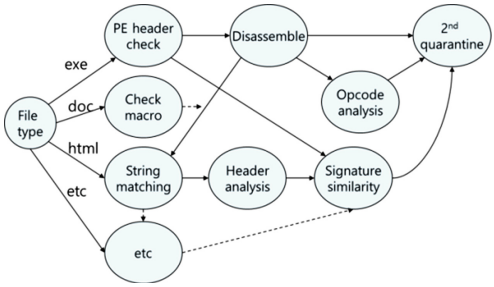


Fig. 3. Graph of static analysis in the primary quarantine [4]

Primary quarantine station has scoring system. Each node raises the scores of the input packet when it determines that the packet is malicious code. In this way, the packet passes several nodes and if score of the packet exceeds the predetermined point, the packet is discarded. The packet has score under the predetermined point when the packet passed final node, it is considered to be a normal packet and transmitted to clean area.

Malicious code classification system using Minhash can be a node in primary quarantine station. After preprocessor reassembles the input packet, Minhash node measures degree of similarity. If it will come out exactly the same, degree of similarity is 100 %. Then it can be determined that the malicious code and immediately discarded. If Minhash value of the input packet is similar to Minhash DB of malicious code, the node gives appropriate point to the packet and passes it to the next node.

The suspicious packet is monitored by Agent. If the process related the packet do the deviated act on the security policy, agent sends events to the manager. Manager may determine the packet to malicious code by analyzing the events. It advertises the packet classified malicious code to Minhash node. Because of this process, it can detect unknown malicious code. Figure 4 shows cyber quarantine system applied malicious code classification system using Minhash.

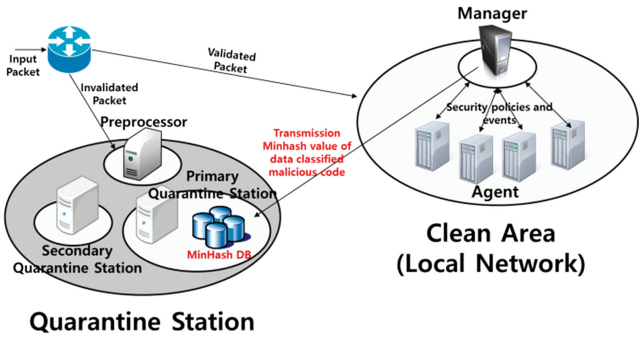


Fig. 4. Cyber quarantine system applied malicious code classification system using Minhash [4]

5 Conclusion

In this paper, we suggest the system that detects malware using Minhash, measures the similarity with existing malicious code and propose to apply this system to cyber quarantine system. Also, this paper verifies the performance of Minhash function. Malicious code classification system using Minhash detected all existing malware. However, it was able to detect only 84 %. It is expected that register a lot of malware on Minhash DB will help increase detection rates.

We will register a lot of malware on Minhash DB, clarify the correlation between the degree of similarity and the scoring system and implement the Minhash node that can be applied to the cyber quarantine system.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0020210).

References

1. Bergeron, J., Debbadi, M., Desharnais, J., Erhioui, M.M., Lavoie, Y., Tawbi, N.: Static detection of malicious code in executable programs (2001)
2. Das, A., Datar, M., Garg, A.: Google news personalization: scalable online collaborative filtering (2007)
3. Han, B., Choi, Y., Bae, B.: Generating malware DNA to classify the similar malwares. J. Korea Inst. Inf. Secur. Cryptology **23**, 679–694 (2013)
4. Lee, S., Chung, T.: Proposal of malicious code classification system using MinHash in network quarantine using SDN. Korea Soc. Digital Ind. Inf. Manage. (2015)
5. Kim, N., Jung, J., Song, Y., Kim, H., Chung, T.: The design of SDN quarantined network. Korea Inf. Process. Soc. (2014)
6. Nath, H.V., Mehtre, B.M.: Static malware analysis using machine learning methods. In: Martínez Pérez, G., Thampi, S.M., Ko, R., Shu, L. (eds.) SNDS 2014. CCIS, vol. 420, pp. 440–450. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54525-2_39

Application of RFID and Computer Vision for the Inventory Management System

Ganjar Alfian[1(✉)], Jaeho Lee[1], Hyejung Ahn[2], and Jongtae Rhee[2]

[1] u-SCM Research Center, Nano Information Technology Academy, Dongguk University-Seoul, Seoul 100-272, Korea
{ganjar, rapidme}@dgu.edu
[2] Department of Industrial and Systems Engineering, Dongguk University-Seoul, 26, 3-ga, Pil-dong, Chung-gu, Seoul, Korea
{macarori, jtrhee}@dgu.edu

Abstract. The RFID technology can be used for item tracking and inventory control. However, the problem such as miss reading and ghost reading usually occur in RFID implementation and has impact on low accuracy of inventory management system. In this study, the computer vision is used to solve the problem of miss reading and ghost reading in RFID. The RFID and computer vision can act as ears and eyes respectively, thus by combining both technologies; it is expected to increase the accuracy of system. The result of experiment has showed that the combination of RFID and computer vision has increased the system accuracy, as the computer vision can help the RFID system to detect the miss reading and ghost reading.

Keywords: RFID · Miss reading · Ghost reading · Computer vision

1 Introduction

RFID (Radio Frequency Identifier) technologies promise to revolutionize future inventory management. They are used to replace the barcode system such that the ID that identifies an object can be accessed wirelessly over a distance [1]. The RFID applications hold tremendous promise, such as RFID tag monitoring. Tag monitoring is a key component in applications like item tracking and inventory control. By frequently scanning its inventory, a retailer can quickly determine if anything is missing, and act accordingly. In addition, the RFID technology has been implemented successfully in Supply Chain Management to reduce the inventory losses [2], for identification and monitoring of agricultural animals [3, 4], food traceability system [5], and healthcare applications to improve patient care as well as reduce overall costs [6].

An RFID passive tag is attached to each item to be monitored. An RFID reader is periodically dispatched to collect all the tags information. A passive RFID tag has limited processing abilities, no power supply, and can only communicate when an RFID reader queries the tag. Since a passive RFID tag is much cheaper than an active tag, passive tags are likely to be deployed on commodity items such as clothing and other retail goods. An active tag is likely to be used on more expensive products.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_92

However, there are at least two problems exist in RFID tag monitoring, such as miss reading and ghost reading. In miss reading, the reader is not able to read the tag in some certain position. This may be due to many reasons such as bad reading angles, blocked signals, or signal dead zones [7]. As opposite, the ghost reading or simply ghost read, is the receipt by the reader of data that is interpreted as valid but was not communicated by a single tag, i.e. the reader receives incorrect data which it interprets as valid data. The identifier received as a result of a ghost read may never have been issued by the numbering authority, may be a valid identifier but not exist on any tag received at the facility experiencing that ghost read, or may exist at that facility but is not in an area where it could possibly be read by the reader experiencing the ghost read [8].

Furthermore, the automatic face analysis which includes, e.g., face detection, face recognition, and facial expression recognition has become a very active topic in computer vision research. Local Binary Pattern (LBP) is one of the most-widely used approach, mainly in face description [9]. In addition, the combination of computer vision (image processing) and RFID technology has been implemented in attendance management system [10], and proved that these two technologies can be implemented together. Thus in this study, the computer vision application and RFID technology will be utilized to solve the miss reading and ghost reading of tag.

2 Methodology

In this study, the application of RFID system in automotive industry is considered as case study. The main function of RFID application in automotive industry is to track/trace the spare part of car. By utilizing this functionality, the certain location of products can be presented in detail from the whole supply chain. For the case of car manufacturing, the spare part of car are placed into special boxes, and special RFID label and tags are attached in boxes. The special RFID tag is presented as the rectangle with the full color, as can be seen in the Fig. 1. In the Fig. 1, the staffs are checking the RFID tag on the box (which the car spare part are placed inside). The color of tag is different with box and presented as bright/contrast color. By designing the tag into

Fig. 1. Special RFID tag, which can easily recognized by computer vision system.

special shape and color, the pattern of tags can be generalized/trained, thus it can be detected by computer vision technique in the last step.

As can be seen in the Fig. 2, the proposed tracking system should recognize the total number of products which are placed on the trolley/cart. The reader is placed nearby the cart, as well as the camera. Both of the RFID reader and camera are connected to the computer. The reader captures the tags which are attached on the product. The total number of tags is counted and presented in the system (computer). In addition, the camera captures every images and send it to system to be recognized by computer vision technique, whether in the image there are products or not (recognized based on the tag pattern and its color). In the last step, the total number of tags from reader are compared with the total number of tags from images, if the result is same, the program/system will show the correct decision, otherwise wrong decision. This computer vision technology is used to double check the tracking system of RFID technology.

In this study, the computer vision technology utilizes LBP (Local Binary Pattern) to recognize the pattern of tags. The LBP (Local Binary Pattern) operator [9] is one of the best performing texture descriptors and it has been widely used in various applications. It has proven to be highly discriminative and its key advantages, namely, its invariance to monotonic gray-level changes and computational efficiency, make it suitable for demanding image analysis tasks. The LBP operator was originally designed for texture description. The operator assigns a label to every pixel of an image by thresholding the 3×3-neighborhood of each pixel with the center pixel value and considering the result as a binary number. Then, the histogram of the labels can be used as a texture descriptor.

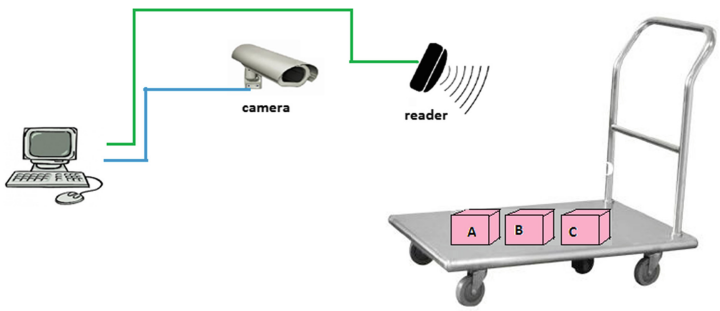


Fig. 2. Integrated RFID and camera for product tracking.

3 Result and Discussion

In this study, the prototype of application is written in Java programming language, and utilize the *openCV* as library which the LBP is included. The *ALR-9900* RFID Reader is used for demonstration while the two IP cameras, *SNV-7084R Samsung* are used to capture the images of the products. The computer *Core i3* is used in this experiment, to receive the data from reader as well as images from IP cameras. The system is

successfully implemented and correctly recognized the number of products which are placed in front of the reader and cameras.

As can be seen in Fig. 3, the boxes which have the special colored tag are placed on the cart or trolley. The total boxes in this experiment are 16 and placed it into two rows, 8 boxes in the each side. For each side, the IP camera is placed around 3 m from the boxes and facing directly to the boxes. The RFID antenna reader is placed near by the products, so it can easily read the tags. As can be seen in Fig. 3, the Graphical User Interface of the system shows two cameras are capturing of the products in each row and the RFID reader is reading the tags of the products. The system calculates the total tags which are read by the reader and by the cameras. The decision of the system is *right*, which mean the number of products which are detected from computer vision technique and RFID tracking system is same/correct.

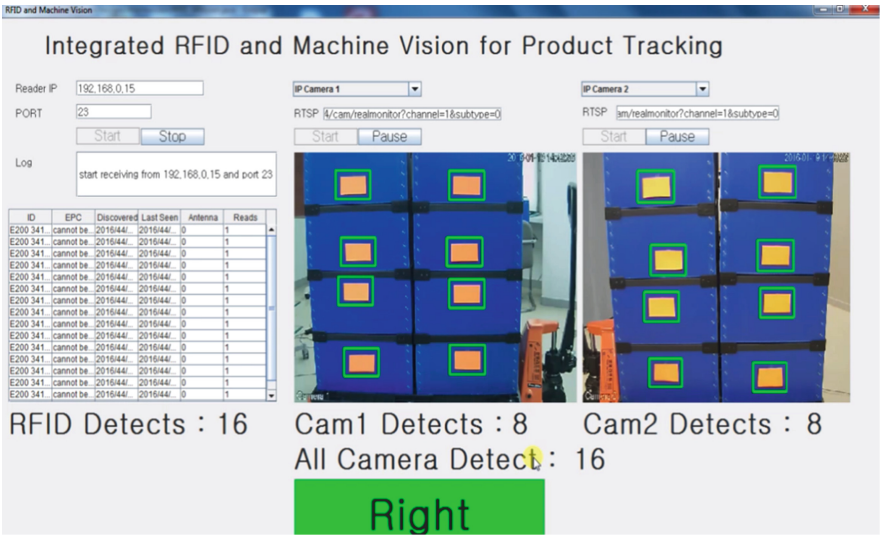


Fig. 3. The RFID reader and computer vision successfully detect number of products.

Furthermore, radio frequencies in reading areas might cause RFID readers to obtain inexistent RFID objects. Thus, the system should be able to avoid this problem. As can be seen from Fig. 4, the boxes are moving on the trolley or cart, while the cameras and reader are capturing the images and tags respectively. In this scenario, the reader is detecting 17 tags, while the computer vision technology is correctly detecting 16 products. The final decision of system is *wrong*, which is the result from RFID reader and computer vision are not match. The RFID reader has showed *wrong* decision, as ghost reading has been occurred in this scenario. In here, we can conclude the computer vision technology is able to avoid the ghost reading.

In addition, RFID readings are noisy in actual situations. This may be due to many reasons such as bad reading angles, blocked signals, or signal dead zones. Miss readings

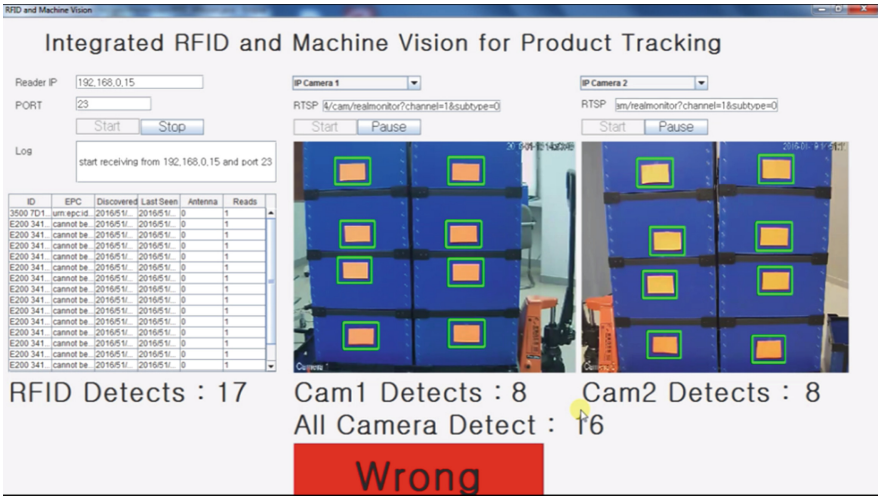


Fig. 4. The different result is detected which is caused by ghost reading.

result in lack of objects' certain positions. As can be seen in Fig. 5, the 16 boxes are moving on the trolley or cart, and reader and cameras are capturing the tag and images respectively at the same time. In this scenario, the reader is capturing 15 tags, while the computer vision technology is correctly detecting 16 boxes which are gathered from two cameras. Thus the final decision is *wrong*, the result from both systems are not correct. In this case, the miss reading has been occurred on RFID reader, while the computer vision system can show real number of products. Thus this computer vision technology is able to avoid the problem of missing tag.

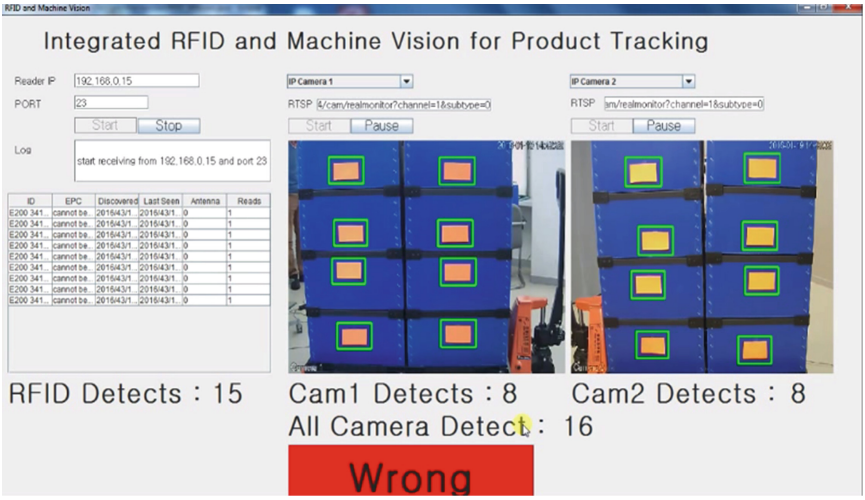


Fig. 5. The different result is detected which is caused by miss reading.

4 Conclusion

In this study, the development of RFID tracking system and computer vision system are presented. The both system are combined together for better decision making, to avoid miss reading and ghost reading which usually occurs in RFID system. Both systems, RFID tracking system and computer vision are developed on Java programming language. The case study on car manufacturing is presented, and prototype of system is developed. The test lab is presented to simulate the real case of car manufacturing. The scenario which involves the products move on the cart is considered. The result of simulation/test lab has showed good result that the computer vision technology can increase the accuracy of system by avoiding miss reading and ghost reading problem which usually occurs in RFID. The computer vision technology can be used to help the RFID tracking system to double check the total number of products, thus it will increase the accuracy of inventory management system.

Future work could be, increasing the number of dataset for training to recognize different object/tag, different angle of images, increasing the camera distance for better decision making and cost analysis of the system.

Acknowledgement. This work was supported in part by the Ministry for Food, Agriculture, Forestry and Fisheries through the Agriculture Research Center under Grant 710003-03-1-SB110 and in part by the 2014 Research Fund through Dongguk University.

References

1. Luo, W., Chen, S., Li, T., Chen, S.: Efficient missing tag detection in RFID systems. In: 2011 Proceedings IEEE INFOCOM, pp. 356–360. IEEE (2011)
2. Sarac, A., Absi, N., Dauzère-Pérès, S.: A literature review on the impact of RFID technologies on supply chain management. Int. J. Prod. Econ. **128**(1), 77–95 (2010)
3. Voulodimos, A.S., Patrikakis, C.Z., Sideridis, A.B., Ntafis, V.A., Xylouri, E.M.: A complete farm management system based on animal identification using RFID technology. Comput. Electron. Agric. **70**(2), 380–388 (2010)
4. Feng, J., Fu, Z., Wang, Z., Xu, M., Zhang, X.: Development and evaluation on a RFID-based traceability system for cattle/beef quality safety in China. Food Control **31**(2), 314–325 (2013)
5. Hong, I-H., Dang Jr., F., Tsai, Y.-H., Liu, C.-S., Lee, W.-T., Wang, M.-L., Chen, P.-C.: An RFID application in the food supply chain: a case study of convenience stores in Taiwan. J. Food Eng. **106**(2), 119–126 (2011)
6. Gaynor, M., Waterman, J.: Design framework for sensors and RFID tags with healthcare applications. Health Policy and Technology, 18 July 2016
7. Xie, D., Xiao, J., Guo, G., Jiang, T.: Processing uncertain RFID data in traceability supply chains. Sci. World J. **2014**, 1–22 (2014)
8. Engels, Daniel W.: On Ghost Reads in RFID Systems. Auto ID white paper (2016). http://cocoa.ethz.ch/downloads/2014/06/None_AUTOIDLABS-WP-SWNET-010.pdf. Accessed 20 June 2016

9. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **28**(12), 2037–2041 (2006)
10. Liu, M., Meng, Q., He, Y.: Research and design the identification software system based on RFID and image processing. In: 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 6, pp. 173–176. IEEE (2010)

Prediction Method for Suspicious Behavior Based on Omni-View Model

Ji-Hyen Choi[1], Jong-Won Choe[1], and Yong-Ik Yoon[1,2(✉)]

[1] Computer Science, Sookmyung Women's University, Seoul, Korea
chl5464@naver.com, {choejn,yiyoon}@sm.ac.kr
[2] Department of IT Engineering, Sookmyung Women's University, Seoul, Korea

Abstract. Recently, CCTV is being applied to prevent crimes. It senses the level of danger as being searching criminal records, a wanted one's montage and so on through mainly Facial recognition. However, it needs additional judging to provide against emergencies, because it cannot predict every criminal situation. In the cause of it, a computer is being fed three-dimensional coordinates from CCTV into a device, and catches not only motion of body or arms but also pattern of hand that were grasped by ConvexHull. And then it predicts suspicious behaviors via judging the movements of an object. Furthermore, to add information about surroundings and location, preventing crimes with more exact judging is the aim on this research.

Keywords: CCTV · Predicting crime · Omni-view · Behavior pattern

1 Introduction

There are more places that installed CCTV to prevent crimes recently. The CCTV has been researched to prevent crimes through facial recognition and prediction of behaviour pattern [1]. Facial recognition consists of extracting facial image that recognized by high-definition CCTV [2]. It is the system which processes making the real-time vigilance possible, if a facial image is identical to a dangerous character that saved in the database. However, it is not every would-be offender belongs to danger list, and it has to assign workers who can watch with vigilance when an object that recognized as dangerous one is seen, until which disappeared. Therefore, the research in respect of an object's behavior prediction is going off. It can assort them two groups: ordinar behaviors or suspicious behaviors, and then analyze recognized objects' gender, distance, motions and so on. But existing development of CCTV technology can grasp motion no more than simplicity, it is the case that cannot grasp motions of small body part like hand or fingers. In its final analysis, behaviour prediction is not commercialization so far. In this paper, supposing it can get over that technological limitations, it predict suspicious behaviors that recognized hand pattern from Kinect camera.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_93

2 Omni-View Model

Figure 1 shows the model to predict suspicious behaviors based on Omni-View. When there are two objects on CCTV, it grasps the distance, body patterns, arm patterns, hand patterns between them in due order [3]. Using Edward T Hall's Intimate distance Table 1 as a criterion [4], the distance between the objects are divided into Public Distance, Personal Distance, Intimate Distance.

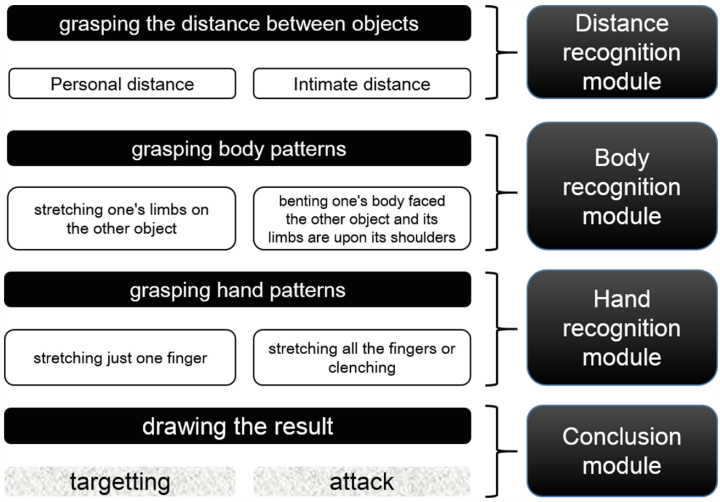


Fig. 1. Model to predict suspicious behaviors based on Omni-View

Table 1. Edward T Hall's Intimate distance

	Distance between objects
Intimate Distance	Less than 45.7 cm
Social Distance	1.2 m–3.7 m
Public Distance	Exceed 3.7 m

2.1 Grasping Body Patterns

Body patterns are grasped through slope of body and motions of limbs. Generally, motion of lower body is steady, and the body is stretched or its upper part is leant during walking. Comparison with it, it judges motions to be suspicious behaviors when the object shows specific motions: leaning its body toward an other object, setting its legs more apart its shoulders, stretching its limbs in the other object, being its limbs upon its shoulders.

2.2 Grasping Hand Patterns

Hand patterns are very important role to judge suspicious behaviors. In the case of violence, the object clenches its fist or stretches all the fingers. And for the targeting direction, the object points at a target with a stretched finger. In the case of targeting, the number of angular points and defects is one or two. For an open hand, it has five angular points and four defects. Of a fist, the number of angular points varies, but the number of defects is set zero.

3 Prediction Algorithm

This paper presumes that is possible to grasp the distance using CCTV because it is existing technology. When CCTV senses the framework of an object, it takes coordinates of some points including joints, head, hands, and feet [5]. The coordinates are stored as 3D XYZ Directions, and it analyses body patterns through making use these.

```
Algorithm Omni View Pattern Cognition
Input : distance between objects, skeleton points
Output : Status( Normal, Warning)

Status = Normal;
IF Social Distance or Family Distance
            THEN Body Pattern Perception();
            return Status;

Body Pattern Perception
Input: skeleton points
Output : Status( Normal, Targeting, Attack)

IF stretching one's limbs on the other object
            THEN Status = Targeting;
ELSE IF Benting one's body faced the other object and its limbs are upon its
shoulders
            THEN Status = Attack;
ELSE
```

```
                THEN Status = Normal;

Hand Pattern Perception(Status);
return Status;

Hand Pattern Perception
Input: Hand points
Output : Status( Normal, warning)

IF Status is Targeting
                THEN IF angular point is 1 or 2, and defect is 1 or 2
                  THEN Status = Warning;
ELSE IF Status is Attack
                THEN IF angular point is 5, and defect is 5
                  THEN Status = Warning;
                ELSE IF defect is 0
                  THEN Status = Warning;
                ELSE THEN Status = Normal;
return Status;
```

Figure 2 shows the model for finding angular points and defects to grasp hand patterns. ConvexHull algorithm uses the system at [6].

It can prevent suspicious behaviors to embody a program based this algorism. Figure 3 shows that is attacking an other object. Like the explained algorism before, it judges that is 'attack' when the distance between objects is less than 45.7 cm, its framework shows whose elbow and shoulder have similar height, the hand and the head have too, and its legs are spreaded like its shoulder width, and the number of apexes is 5 and of defects is 4, or the number of defects is 0. If it is shown the correct 'attack' figure synthesizing the patterns of distance, skeleton, and hands, we can prevent crimes by so alerting warning sound or sending notice message to nearby local branch of police.

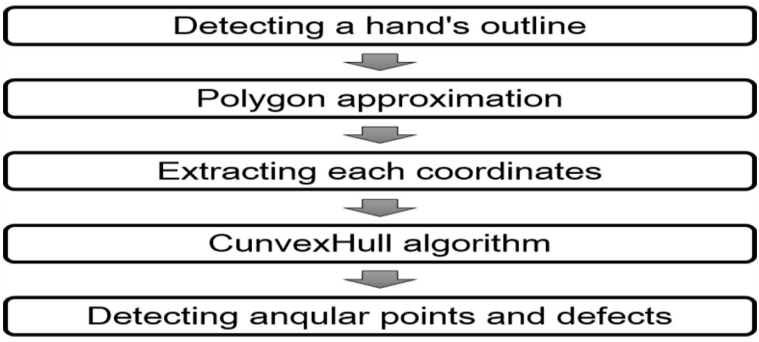


Fig. 2. The model for grasping hand patterns

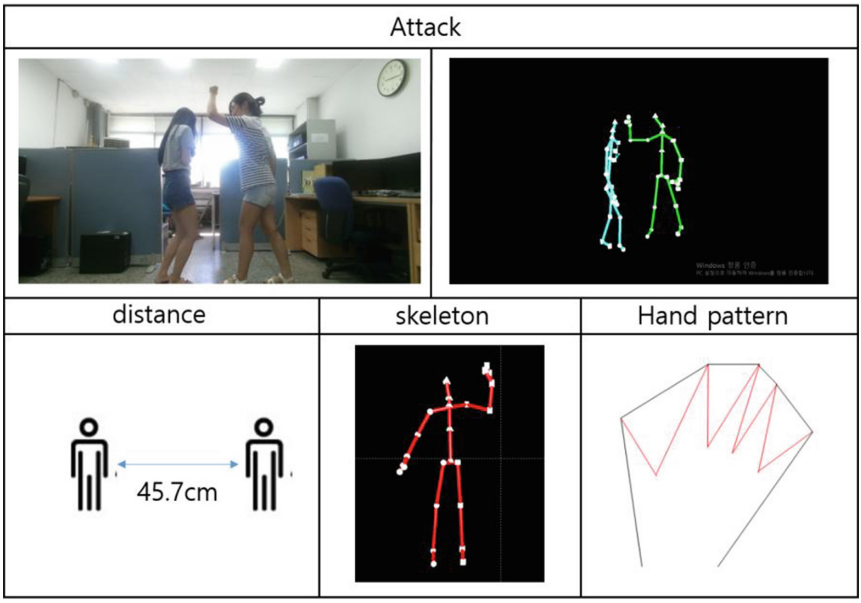


Fig. 3. The results of predediction for Omni-View model for attack behavior. The figure shows attack action like real picture, skeleton model, distance, attack pattern by using skeleton model, and hand pattern by using convexhull.

4 Conclusion

The end of this paper is preventing crimes from arising through predicting and grasping the motions. It can do it if the system that CCTV senses suspicious behaviors by analysing the 3D coordinates of objects' is built, without a worker. If this technology is in common, it becomes easier sensing and preventing crimes without consuming personal resource always. Preventing suspicious behaviors is formed of analysing distance and body, arm, hand patterns in due. When it applies the technology, also be applied facial recognition as well as VMS vigilance to existing CCTV, it can make more advanced preventing crimes possible. In addition to these, it is expected to sense more exactly if it can make analysing continuous motions possible.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R2718-16-0004) supervised by the IITP (National IT Industry Promotion Agency).

References

1. Jung, S.-H., Cha, J.-S.: Study on CCTV based sensor applied SURF image processin system. Korea Sci. Art Forum **16**, 403–408 (2014)
2. Korea Internet & Security Agency. A Study on Security Framework for CCTV based Face Detection and Recognition System
3. Choi, J.-H., Choe, J.-W., Yoon, Y.-I.: Crime prediction model with moving behavior pattern
4. Hall, E.T.: Proxemics
5. Webb, J., Ashley, J.: Beginning kinect programming with the microsoft kinect SDK, bjpublic
6. Park, B.-J., Lee, J.-H.: An improved convex hull algorithm considering sort in plane point set. J. IKEEE **17**(1), 29–35 (2013)

Optimal 3D Printing Direction for Stability of Slanted Shapes

Jiyoung Park(✉) and Hwa Seon Shin

Korea Electronics Technology Institute, Daewangpangyo-ro 712 Beon-gil, Bundang-gu, Seongnam-si, Gyeonggi-do 463-400, Korea
{jyp,L544}@keti.re.kr

Abstract. The Fused Deposition Method (FDM) constructs objects in layers of melted material. Because the layer-by-layer construction of slanted shapes introduces a smaller bonding area between layers and significant warping, users often heuristically optimize the printing direction. We propose a novel method for selecting the optimal 3D printing direction to increase the stability of slanted shapes. Our optimal direction leads slanted subparts to be oriented either perpendicular or parallel to the build plate as much as possible. We first find a set of stable directions for the input model and then validate each of them using the following three criteria: similarity to dominant directions, stability and material cost. The experiments show that our optimal printing direction enables the given shape to be printed without undesired deformation during printing process.

Keywords: Fused deposition method · 3D printing direction · Orientation optimization

1 Introduction

The Fused Deposition Method (FDM) is a major 3D printing technique that constructs objects in layers. Thermoplastic material is melted and extruded through a heated printer head nozzle. Because FDM accumulates layers in a certain printing direction, printing slanted shapes creates a smaller bonding area between layers and uneven warping during cool down. Moreover, the forces applied by the print head on the upper levels generate torque on the part, which bends the structure, and the resulting deformation can lead to print failures in the worst case scenario [1]. A common solution is to optimize the printing direction. Before printing, users rotate the input model such that sloping subparts become close to perpendicular or parallel to the build plate. However, finding an optimal direction often requires multiple trial-and-error iterations, which waste time and printing material. Although many methods have been proposed for the automatic computation of optimal printing directions, none of them consider how to avoid the undesired deformation of slanted shapes during the printing process.

We present a novel method of finding the optimal printing direction to increase the stability of slanted shapes. In practice, deformation arises from many factors (e.g., material, temperature, layer thickness), but a physical simulation incorporating the temperature and dynamic forces of the entire printing process is not trivial.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_94

Therefore, we use a geometrical approach inspired by the observation that vertical and horizontal shapes are more reliable than slanted shapes. Our optimal direction is determined by a set of dominant directions that best describes the variance in the slope angles of the slanted subparts. Given a cylinder, for example, the three dominant directions are defined by the cylinder's height axis and the two perpendicular axes of the base ellipse. Printing with a dominant direction enables many slanted subparts to be oriented either vertically or horizontally on the build plate. Our solution can avoid undesired deformation in the printing process without complex physical simulation. We further consider static stability and material cost for printing, formulated as a combination of the three metrics.

2 Related Work

Printing direction is one of the primary factors that alter the resulting quality and time cost of the printed part. A number of approaches optimize the printing direction with a variety of objectives, for example, to improve surface quality, decrease material, and reduce printing time. The most common methodology is the multi-criteria assessment approach [1, 2]. It begins with a set of candidate directions as a reduced search space. The candidates are then evaluated by an assessment function of multiple attributes such as time cost, material cost and surface error. The weighted sum of a set of the attribute scores results in the final optimal direction.

In the recent literature, efforts have been made to tackle FDM's drawbacks in terms of appearance and stability. Zhang et al. [3] focused on the surface artifacts on the area contacted by supporting structures. They optimized printing direction to avoid placing supports in perceptually significant regions. The scores of a model in a range of orientations are evaluated by a combination of support area, visual saliency, preferred viewpoints, and smoothness preservation. Umetani and Schmidt [4] proposed an optimal printing direction to increase the model strength after printing. However, no approaches have addressed the shape deformation problem during the printing process. Recently, Dumas et al. [5] highlighted the low reliability of slanted pillars in tree-like support structures. They proposed a robust support structure based on vertical pillars and horizontal bars.

3 Method

Our optimal printing direction is based on the following three criteria: dominant directions, static stability, and support material cost. Figure 1 shows our pipeline. We begin with the input model and search for a set of candidate directions Φ as a search space for the optimal direction. We then obtain the final optimal direction ϕ_{opt} as

$$\phi_{opt} = \underset{\phi \in \Phi, |\phi|=1}{\arg\max} (D_\phi + A_\phi + V_\phi), \tag{1}$$

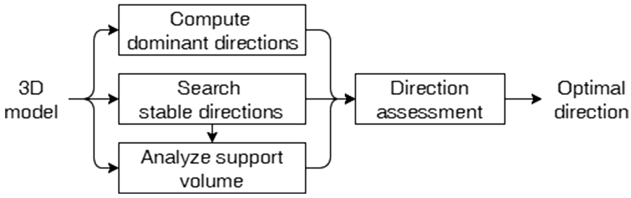


Fig. 1. Outline of our method.

where D_ϕ is the measure of how close a candidate direction ϕ is to the dominant directions, A_ϕ is the size of the contact area created by the model printed in the direction ϕ, and V_ϕ is the support volume built beneath the overhang area of the model.

Dominant Directions. Given the input model, we find a set of dominant directions $\mathbf{e}_{i\in\{0,\ldots,5\}}$ based on the normal clustering technique [6, 7]. We start with all facet normals and classify them into three groups based on the likeness to the cluster center. For each group, we find the directional center using PCA. Then, we apply the SVD to compute the closest three orthogonal directions that are the new cluster centers. We repeat the three steps until convergence. The three initial cluster centers are (1,0,0), (0,1,0), and (0,0,1). The resulting three orthogonal directions are the first three dominant directions $\mathbf{e}_{i\in\{0,1,2\}}$, and the other three are defined by $\mathbf{e}_{i\in\{3,4,5\}} = -\mathbf{e}_{i-3}$. We finally estimate how close a candidate direction ϕ is to the dominant directions by

$$D_\phi = \max_{\phi\in\Phi,|\phi|=1,i\in\{0,\ldots,5\}} (\phi \cdot \mathbf{e}_i). \tag{2}$$

Stable Directions. We use the static stability criterion to find a set of candidate directions as well as to assess each candidate for the final selection of the optimal direction. The static stability is related to the proportion of the mass of the object located above the base [8]. We estimate two geometric attributes: the position of the center of mass projected onto the convex hull and the size of base polygon. The first attribute finds the candidate directions Φ, while the second determines A_ϕ of the assessment function. First, we construct the 3D convex hull of the given model. We then project the center of mass in each facet normal direction and compute the intersection. If the hit position is inside the facet, the normal direction is stable and is chosen as a candidate for the final selection. A base polygon is defined by the facets of the convex hull that are in contact with the build plate. Initially, we specify a base polygon for the candidate direction ϕ as a set $\Omega_\phi = \{t_\phi\}$, where t_ϕ indicates the facet whose normal direction is ϕ. We then expand Ω_ϕ by connecting its neighboring facets by examining how close the facet normal is to ϕ, which is written as

$$\Omega_\phi = \Omega_\phi \cup \{t\}, \quad \text{if} \quad \phi \cdot \mathbf{n}_t > \varepsilon,\ t \in adj(\Omega_\phi), \tag{3}$$

where $t \in adj(\Omega_\phi)$ denotes facets incident to the base polygon, and ε is a predefined threshold. The areas of the facets in Ω_ϕ are summed up, and A_ϕ is computed by

$$A_\phi = \sum_{t \in \Omega_\phi} Area(t) \Big/ \sum_{\omega \in \mathrm{M}} Area(\omega), \tag{4}$$

where M is the input surface model. If no stable facets are found, the candidates are defined as the facets whose normals are closest to the dominant direction.

Support Volume. Printing overhanging parts requires a disposable support structure. A well-known support structure is a volumetric shape built by extruding overhang facets downward. As the overhanging part varies with the printing direction, the support volume can validate the direction's cost efficiency. Accordingly, we employ a support volume to minimize the printing cost. We find overhang facets by testing whether the angle between the plane defined by the facet and the candidate direction is higher than a critical angle. The downward extrusion of a facet can approximate a trigonal pillar. The support volume is defined as the aggregate of all the trigonal pillars. Once the model is oriented in candidate direction ϕ, the resulting measure for the support volume is finally computed by

$$V_\phi = \sum_{t \in \{overhang\}} Area(t) \cos\theta_\phi d \Big/ Vol_{bbox}, \tag{5}$$

where θ_ϕ is the angle between the facet t's normal vector and ϕ, and d is the distance between the center of mass and the build plate ($z = 0$). Note that we use the volume of the model's bounding box Vol_{bbox} for normalization.

4 Result

We 3D printed several models in two different directions using Ultimaker 2 and Cura. We implemented our software using Libigl[1] (for geometric processing) and Qhull[2] (for convex hull computation).

Figure 2 shows the printed Giraffe[3] model. The printing material was Wood fill produced by ColorFabb[4]. We used a 0.6 mm nozzle and a layer thickness of 0.12 mm. The two printing directions and the printing results are shown in Fig. 2(a) and (b). The first direction could reduce overhang area on the model surface. However, it resulted in serious deformation of the four legs. In contrast, printing with our optimal direction constructed the input model correctly.

Figure 3 shows the Male Javelin[5] model printed with a 0.4 mm nozzle and a layer thickness of 0.12 mm. The printing material was ColorFabb's XT. Our optimal

[1] http://libigl.github.io/libigl/.

[2] http://www.qhull.org/.

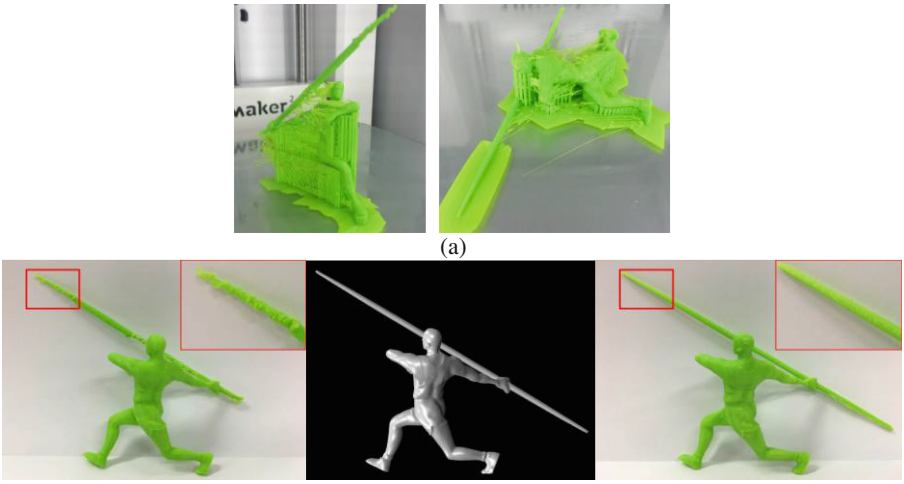
[3] http://www.123dapp.com/123C-3D-Model/Giraffe/593790.

[4] http://colorfabb.com/.

[5] http://www.123dapp.com/123C-3D-Model/Male-Javelin/719596.

(a) (b)

Fig. 2. The Giraffe model printed in two different directions. The direction shown in the top of (a) avoided a large overhang area but resulted in serious deformation in the legs ((b), left) compared to the original model ((b), middle). In contrast, our optimal direction laid the model on the build plate ((a), bottom). The right image in (b) shows that the model stands upright, just like the original. The print time and material usage for our direction were 7 h 46 and 8.40 m, respectively, as opposed to 7 h 05 and 8.26 m for the other direction.



(a)

(b)

Fig. 3. The Male Javelin model printed in two different directions. The left picture in (b) shows the model printed in the default direction. The print time and material usage were 3 h 27 and 1.73 m. The model printed in our optimal direction ((b), right) shows much better surface quality with less time (3 h 08 and 1.79 m) particularly in the upper part of the javelin.

direction is shown in the right image of Fig. 3(a), and the printing result is shown in the right image of Fig. 3(b). Our optimal direction resulted in better surface quality, particularly in the javelin. Our optimal direction made the javelin parallel to the build plate, while the other direction inevitably created aliasing to construct the slope. Moreover, our printing direction allowed the lower part of the printed javelin to be well

printed. We observed that printing software tools often failed to build supporting geometry in spike shapes, such as the lower part of the printed javelin. We tested two different software tools, Cura and Simplify3D, and both missed the same part without the direction optimization.

5 Conclusion

We have shown how to optimize the printing direction to avoid shape deformation during 3D printing. Our experimental results showed that our optimal printing direction enabled thin slanted shapes to be printed correctly without undesired deformation or partial loss in subparts such as spike shapes. Moreover, our stability criterion allows the object to stand stably during the printing process. Further concerns include the support volume and achieving reasonable printing time and material costs.

Our normal clustering-based dominant directions, however, cannot effectively represent radial shapes composed of many spikes. For future research, it is worth studying how to select important subparts with high potential for deformation in automated or interactive ways.

Acknowledgments. This work was supported by the IT R&D program of MSIP/IITP. [R0126-15-1016, Custom smart slicer development for domestic popular 3D printer].

References

1. Xu, F., Wong, Y.S., Loh, H.T., Fuh, J.Y.H., Miyazawa, T.: Optimal orientation with variable slicing in stereolithography. Rapid Prototyping J. **3**(3), 76–88 (1997)
2. Pandey, P.M., Venkata Reddy, N., Dhande, S.G.: Part deposition orientation studies in layered manufacturing. J. Mater. Process. Technol. **185**(1–3), 125–131 (2007)
3. Zhang, X., Le, X., Panotopoulou, A., Whiting, E., Wang, C.C.L.: Perceptual models of preference in 3D printing direction. ACM Trans. Graph. **34**(6), 215:1–215:12 (2015)
4. Umetani, N., Schmidt, R.: Cross-sectional structural analysis for 3D printing optimization. In: SIGGRAPH Asia 2013 Technical Briefs, pp. 5:1–5:4 (2013)
5. Dumas, J., Hergel, J., Lefebvre, S.: Bridging the gap: automated steady scaffoldings for 3D printing. ACM Trans. Graph. **33**(4), 98:1–98:10 (2014)
6. Hildebrand, K., Bickel, B., Alexa, M.: Orthogonal slicing for additive manufacturing. Comput. Graph. **37**(6), 669–675 (2013)
7. Reisner-Kollmann, I., Luksch, C., Schwärzler, M.: Reconstructing buildings as textured low poly meshes from point clouds and images. In: Eurographics Short Papers, pp. 17–20, (2011)
8. Fu, H., Cohen-Or, D., Dror, G., Sheffer, A.: Upright orientation of man-made objects. ACM Trans. Graph. **27**(3), 42:1–42:7 (2008)

A Study on DDS-Based BLE Profile Adaptor for Solving BLE Data Heterogeneity in Internet of Things

Jung-Hoon Oh[1], Moon-Ki Back[1], Gil-Tak Oh[1], and Kyu-Chul Lee[2(✉)]

[1] Department of Computer Engineering, Chungnam National University, 220 Gung-Dong, Yuseong-gu, Daejeon, Korea
{vicers5, zmzment, ogt0329}@cnu.ac.kr
[2] Department of Computer Engineering, Chungnam National University, 99, Daehak-ro, Daejeon, Korea
kclee@cnu.ac.kr

Abstract. For communication between heterogeneous objects there is a heterogeneity problem that needs to be solved and to do this interoperability between different protocols has to be secured using a middleware structured adaptor. In this paper we suggest using a BLE(Bluetooth Low Energy) profile adaptor which Interoperates data between objects based on DDS(Data Distribution Service), a real time standard middleware. With this BLE profile adaptor, BLE devices and other protocoled devices can interoperate data and by using the profile based Standard data format we can obtain wide interoperability between devices regardless of the BLE device's type or manufacturer.

Keywords: IoT(Internet of Things) · BLE(Bluetooth Low Energy) · DDS(Data Distribution Service) · Inter-operation · Profile · Adaptor · Heterogeneity

1 Introduction

With the indefinite development in IoT, the existing internet environment is extending to distributed connection between devices [1]. These devices are using appropriate protocols for communication given the physical properties and areas of activity. In this process there has been a heterogeneity problem in communications between different protocols. To solving this heterogeneity problem, securing interoperability is crucial and to do this, a middle-ware based internet of things architecture is needed [2]. In this paper, we discuss about BLE profile adaptors in IoT environments for securing interoperability. BLE profile adaptors solve the heterogeneity problem between BLE and other protocols letting us use the BLE data in other protocol fields which gives us a much wider appliance. Also, because many researches and commonly used devices use an individually defined data format, there is a heterogeneity problem where the different BLE devices depend on the type and manufacturer to be compatible with each other. For the solution of this, the BLE profile adaptor defines and applies a profile [3] based sectoral common data format, a standard data format, to solve the heterogeneity

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_95

problem between BLE devices giving the users a wide interoperability between BLE devices regardless of the type or manufacturer.

The remainder of this paper is organized as follows. Section 2 introduces a research on securing interoperability between objects. In Sect. 3, we describe BLE profile application methods along with BLE profile adaptor structures and descriptions on the interlocking process. Lastly, we discuss conclusion of our findings and further research directions.

2 Related Works

[4] is a mobile gateway research for ubiquitous mobile health scenario. The healthcare device gathers information from the users and integrates it with the smart device's location information and then delivers the integrated information to the cloud. Based on the delivered information, the device finds and delivers a suitable service for the user. [5] gives the user services in home automation environment based on the user's location. We matrix map the information gathered by the device and the information of the user's location to find a suitable service for the user and use CoAP(Constrained Application Protocol) to deliver the service. [6] gathers information from the device and by using semantic web, it offers the user and the cloud services through Restful. In the perspective of scalability, [4–6] has a disadvantage when the nodes used for 1:1 communication external links increases, the complexity rapidly increases, which leads the scalability to decrease. On the other hand, the BLE profile adaptor uses DDS [7] carrying mesh topology type Pub/Sub structure where every node splits the burden regardless of the number of nodes, giving the device great scalability. In the perspective of profile use, [4–6] don't use profile, the standard data format, which makes a heterogeneity problem between BLE devices even though the data has the same meaning. The BLE profile adaptor however, uses the standard data format profile which gives the device a wide interoperability regardless of the type or manufacturer.

3 Research Content

3.1 BLE Profile Application

The BLE profile has a standard data format constituted by the Bluetooth SIG consisting of many types of different profiles with different functions [8]. Each profile has services, their characteristics, their characteristic's requirements, etc. Many researches and common devices in [4–6] don't use profile and uses individual data format. This causes heterogeneity problems between BLE devices. To solve this problem, we apply BLE profile to the BLE adapter. To do this, we classify profiles by their fields based their functions. After, we extract the common parts of the fields we classified, make a common data format using the extracted parts and apply them to the BLE profile adaptor. We divided 12 service profiles into 4 fields. We used each profile's characteristics and their requirements to extract the common data parts. Each characteristic's requirements were divided into 3 groups, Mandatory, Optional, and Conditional for

Profile classification	Profile Type	Each Profile IDL Format

Health	• Blood Pressure
	• Glucose
	• Health Thermometer
	• Heart Rate
		Struct Health_Topic{
		Short Profile Type;
		Byte[14] Measurement;
		Short Feature;
		Float RACP;
		}

Exercise	• Cycling Power
	• Cycling Speed and Cadence
	• Running Speed and Cadence
		Struct Exercise_Topic{
		Short Profile Type;
		Byte[5] Measurement;
		Float Feature;
		Byte[1] Location;
		}

Alert	• Alert Notification
	• Phone Alert Status
	• Find Me
		Struct Alert_Topic{
		Short Profile Type;
		Float Alert;
		Short Status;
		Short Control;
		}

Location	• Location and Navigation
	• Proximity
		Struct Loacation_Topic{
		Short Profile Type;
		Float Feature;
		Byte[13] Location;
		Long Navigation;
		}

Fig. 1. Profile type & each profile IDL format

each characteristic. And we extracted the Mandatory element of each characteristic and made a common data format between the profile's fields as shown in Fig. 1. We applied the common data format to the BLE profile adaptor so that the adaptor can use the standard profile based common data for each sector. By doing this, we used the BLE profile adaptor to solve the heterogeneity problem between BLE devices and can deliver wide interoperability regardless of the type and manufacturer of the BLE device.

3.2 BLE Profile Adaptor's Structure

The structure of the BLE profile adaptor is as shown in Fig. 2. The BLE profile adaptor is BLE 4.0 in standard with 5 components and BLE zones. The BLE zone is constructed with the BLE device's Peripheral and Central components and the BLE profile adaptor's Peripheral, Central components. When the Central requests pairing, the Peripheral data is sent. But compared to the Central component, which can have several connections, the Peripheral, can only have one connection. For this drawback of the Peripheral component, we increased the number of them to 3 in the BLE profile adaptor. In the BLE profile adaptor's 5 components there is a Device Management part for the Central component and 3 Peripheral components, Device Abstraction which abstracts each of the connected device's data from the received BLE data, Profile Abstraction, which converts the abstracted data into sectoral common data as shown in Fig. 1, Topic Instance Manager, which creates Topic by sectoral common data and converting the Topic into sectoral common data, and finally the DDS Entity Manager

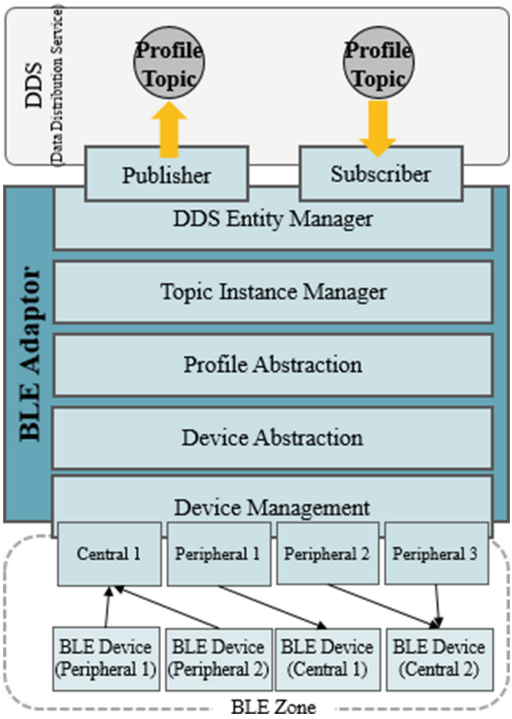


Fig. 2. BLE profile adaptor

which (publish/subscribe)s the topics and (generate/manage/delete/detect)s DDS related Entity.

3.3 BLE Profile Adaptor Operating Procedure

The operating procedure of the BLE profile adaptor can be divided into the publishing procedure and the subscription procedure of the DDS. In the publishing procedure, the initial procedure of the BLE profile adaptor is connected with the BLE devices. The Central component of the BLE profile adaptor requests connection to the BLE devices (Peripheral) and if the Peripheral has data to transmit, the connection is concluded. After the connection is made, the Peripheral components transmits BLE Packets to the BLE profile adaptor, extracts necessary data, and delivers the data do Device Abstraction. Device Management only has a function of managing the Central and Peripheral components of the BLE profile adaptor. Device Management not only functions as constantly requesting for connection of BLE Devices but also instructs to sever connections forged. Data transmitted to Device Abstraction is managed and abstracted by the connected devices(Peripheral). After the process of data abstraction done by the devices, we can clearly identify the device's ID, MAC Address, Data Profile Type, Data Value, and Connection State. After that, the abstracted data is

delivered to Profile Abstraction and is abstracted into the common data format for each devices that's based on the research contents shown in Fig. 1. The common data format for each device is generated into 4 types of Topics from the Topic Instance Manager and through the DDS Entity Manager's Publisher, is published to the DDS.

For the subscription process, the BLE profile adaptor subscribes a Topic from the DDS through the DDS Entity Manager's Subscriber. The subscribed Topic that is primarily validated from the Topic Instance Manager is converted into each device's common profile data. The converted data is delivered to Profile Abstraction and is once again converted into each device's common data, the BLE profile adaptor identifies the Device's ID, MAC Address, Data Profile Type, Data Value, Connection State, and Etc. The common data for each device goes through Device Abstraction and Peripheral components and becomes a BLE Packet. After the process, Device Management concludes necessary connections with the BLE device(Central) and through the BLE profile adaptor's Peripheral component, is transmitted to BLE Packet.

4 Conclusion and Further Research Directions

To provide wide use of the BLE protocols in the environment of IoT, we have to solve the heterogeneity problem by securing interoperability of different protocols. Also, in the case of BLE, even though there is profile, a standard data format, the use of individually defined data format causes heterogeneity problems between BLE devices. In this paper, we have studied BLE adapters based on DDS for solving the heterogeneity problem that hinder interoperability between standard profiles. Having a BLE profile based sectoral common data format applied to the BLE profile adaptor we made it so that the adaptor is not affected by type or manufacturer nor does it cause heterogeneity problems between the devices, giving the device a wide interoperability.

In this study, there is a problem where we can't apply a presently active extension known as Beacon. We are thinking of setting our research direction towards combining the BLE profile adaptor with Beacon and as a result, creating an area based BLE profile adaptor unifying BLE data with proximity information.

Acknowledgements. This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1066721).

References

1. Elkhodr, M., Shahrestani, S., Cheung, H.: The internet of things: new interoperability, management and security challenges. Intl. J. Netw. Secur. Its Appl. (IJNSA) **8**(2), 85–102 (2016)
2. Oh, J.-H., Back, M.-K., Lee, K.-C.: DDS-Based beacon adaptor for real-time proximity interoperation in internet of things. Database Res. **31**(3), 14–26 (2015)
3. Bluetooth Special Interest Group: Profile (2016). https://developer.bluetooth.org

4. Santos, J., Silva, B.M.C., Rodrigues, J.J.P.C., Casal, J., Saleem, K.: Internet of things mobile gateway services for intelligent personal assistants. In: 17th International Conference on E-health Networking, Application & Services (HealthCom), pp. 311–316 (2015)
5. Mainetti, L., Mighali, V., Patrono, L.: A location-aware architecture for heterogeneous building automation systems. In: International Symposium on Integrated Network Management, pp. 1065–1070, May 2015
6. Datta, S.K., Bonnet, C., Gyrard, A., Ferreira da Costa, R.P., Boudaoud, K.: Applying internet of things for personalized healthcare in smart homes. In: Wireless and Optical Communication Conference, pp. 164–169, Oct 2015
7. OMG: Data Distribution Service for Real-time Systems version 1.2, January 2007. http://www.omg.org/spec/DDS/1.2/PDF/
8. Bluetooth Special Interest Group: Profile (2016). https://developer.bluetooth.org

A Study of Environment-Adaptive Intrusion Detection System

Ki-Hyun Lee[1] and Young B. Park[2(✉)]

[1] Department of Computer Engineering, Dankook University, Juk-Jeon, Republic of Korea
qkqnl47@gmail.com
[2] Department of Computer Science, Dankook University, Cheon-An, Republic of Korea
ybpark@dankook.ac.kr

Abstract. Recently, the intrusion cases by hackers are growing fast. In order to prevent such intrusions, it is common to install a firewall or an intrusion detecting system to be employed. However, since traditional intrusion detecting system detects attacks by using static attack signatures, there are limits in coping with changes of environment or methods of sneaking in. For these problems to be solved, this paper suggests Environment Adaptive Intrusion Detection System using MAPE model. The system identifies the changes of external environment from MAPE's acts; through the recognized change values of environment, it creates a proper attack sign and applies to the intrusion detecting system. It is expected to operate the Environment Adaptive Intrusion Detection System by using this method.

Keywords: Intrusion detection system · Environment adaptive · MAPE

1 Introduction

Due to developed IT technologies and the advent of new technologies, hackers' attacking routes have been complicated and sophisticated. Through such evolved intrusion methods, loads of security incidents have been occurring annually [1]. In methods of prevention those attack, People have installed traditional firewall in internal, external system, but it was not enough to protect the system from the sophisticated intrusions. To solve these matters, there various IDSs (Intrusion Detection System) come in to being [2, 3].

IDS is operated by pre-defined rule (Attack signature), according to the defined rule, the performance of intrusion detection is settled. For this reason, the rule is very important in IDS. In general IDS, men create the rule manually, but there are many difficulties for men to analyze and write the code manually. In order to overcome these difficulties, IDS using a variety of automatic rule generation methods are suggested [4–6]. By using suggested methods above, we have been able to operate IDS effectively, but there are finite limits which consider changes of the deployed environment of IDS. Such limits could generate problems like False-Positive, False-Negative. To solve these, this paper suggests an IDS which uses MAPE model to adapt changes of environment.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_96

2 Background

2.1 IDS (Intrusion Detection System)

IDS is a system to detect intrusions done by attackers, and to defend these actions actively. This system consists of Abuse Detection Method and Anomaly Detection Method, generally it uses the former. Also, in terms of the locations the system is deployed, there are two things to be classified as NIDS (Network Intrusion Detection System and HIDS (Host-based Intrusion Detection System). Such systems are decided the detection performance by which the users define the rules, so the research in enhancing the performance is under processing [7, 8].

2.2 MAPE Model

MAPE is a model to cope with actively changing environment and is being used for a research of self-adaptive systems [9]. Such MAPE model largely consists of four modules, and these are Monitor, Analysis, Plan, and Execute models. Each module plays a role. Monitor monitors a dynamical change; Analysis analyzes detected changes or characteristics of data; Plan makes a solution to adapt to the changed environment; Execute plays a role to actually carry out the scripted solution. These four modules consistently, repeatedly continue executing.

3 Environment-Adaptive Intrusion Detection System

The Environment Adaptive IDS that the paper suggested has a structure in which intrusion detection module is connected with MAPE model. Through the MAPE model, monitors the environment changes, and if any changes are detected, create a rule which adapted to environment change. After that, created rule is applied the rule to a detection engine of IDS. Figure 1 shows the architecture of suggested IDS.

3.1 IDS Module

IDS module uses snort which is an existing network intrusion system [2]. NIDS is operated by rule based on rule, analyses network packets and carries out tasks such as detecting packets, logging in, and giving alerts. It is based on rule, according to the rule applied a new shape of IDS can be implemented. In this paper, the MAPE model is in charge of the renewal work of the rule to implement the Environment Adaptive IDS.

3.2 Adaptive Module

The role of Adaptive module is to construct a rule that is adapted to environment according to changes in the environment. In order to make it possible to the system to be adaptive to environment, the adaptive module utilizes MAPE Adaptive Framework.

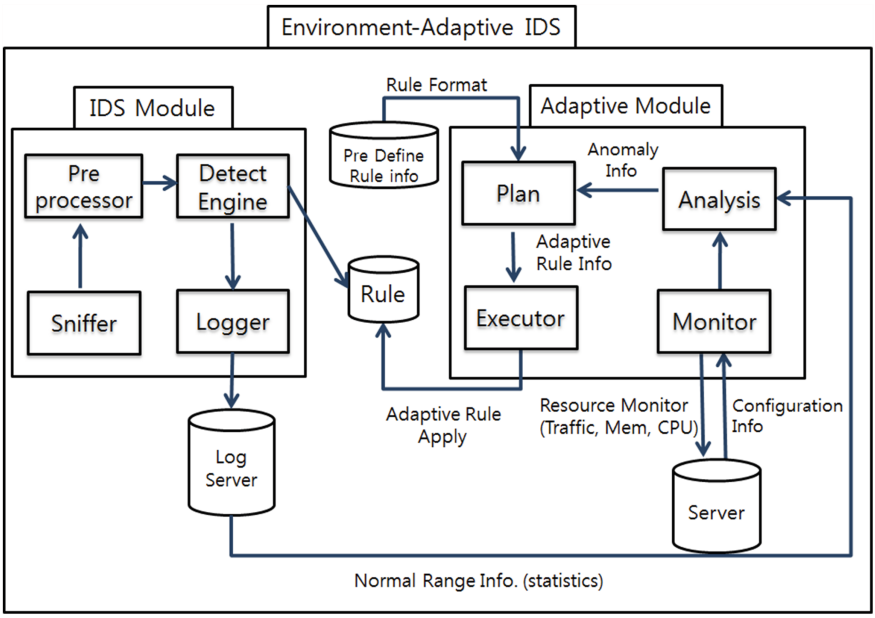


Fig. 1. The architecture of suggested IDS

MAPE Framework consists of Monitor, Analysis, Plan and Execute, and each role is explained in the following.

3.2.1 Monitor Module

Monitor module is a module that plays a role to consistently monitor the environment changes. The values that Monitor module observes are network traffic, usage of server memory or the amount of CPU usage, etc. Within the Monitor module, there are several probes to be deployed to monitor the various changes of environment values at the same time. Also, in Monitor module, it designates a threshold value, so if any environmental changes occurred exceed the designated threshold value, the monitored environmental change values are sent to the Analysis module. For the reason why it uses the threshold value in Monitor Module is when every time it occurs and creates a rule, it might drop the performance.

3.2.2 Analysis Module

In analysis module, it identifies an abnormal environmental change, delivers a decision of whether a new Rule should be made or not to the Plan module. The abnormal identification of environmental change uses a statistic method. The reason why it uses the statistic method is it needs a normal model of environmental change in order to decide whether environmental change is normal or not. Normal model of environmental change is expressed in a range form, has a constant range in the criterion of mean value collected from the past. This normal environmental change model could have a different model in terms of scale of environment to be deployed or

characteristics. By using this normal model of environmental change, Analysis module decides all the environmental changes which are not in the normal range, delivers environmental change information to Plan module.

3.2.3 Plan Module

In Plan module, it generates a rule that considers an environmental change value. However, considering environmental change values, generating a rule is a very tough work. For the reason, in this paper, environment adaptive rule are created by using pre-defined rule format and analyzed environment change value. Figure 2 shows the Example of Pre-Defined Rule Format.

```
alert tcp any any -> any any (msg" Syn Flooding"; flag: S; flow stateless;
threshold: type both, track by_src, count X, secdonds Y )
```

Fig. 2. Example of pre-defined rule format

In Fig. 2, X and Y are environmental variables. Environment adaptive rule are constructed by applying environment changed variables which is X and Y. When generating rules, it can generate different rules by applying different environmental values even though it uses same format.

3.2.4 Execute Module

Execute model applies the rule received from Plan module to IDS module. Execute module converts rule format into IDS module compatible format in order to renew the IDS rule. After that, by applying the rule format to IDS rule, it could carry out the intrusion detection using the environmentally adapted rule.

4 Evaluation

To evaluate a proposed method, the detection ratio of traditional IDS and proposed adaptive IDS against DoS(Denial of Service Attack) are compared at the environment changes on server In the evaluation, the hping3 tool is used to generate DoS attack.

Also, following assumption is used to simulate;

Free state of server: 300 packet per second
Free state of server: 600 packet per second
Free state of server: 900 packet per second

4.1 Scenario 1 Server State Changes from Free to Busy

In order to measure detection ratio between traditional IDS and proposed adaptive IDS, simulated Dos attack is conducted. For the 100 ~ 1000 packets per second, 100 DoS

Table 1. Result of simulation in Server state: Free -> Busy

	Traditional IDS	Adaptive IDS
True-Positive	48	83
True-Negative	25	17
False-Positive	0	0
False-Negative	27	0

attacks are conducted to simulate situation when, state of server changes from free to busy, Traditional IDS used static rules that are created in normal state, and proposed adaptive IDS used environment-adaptive rules.

Table 1 shows result of simulation in state of server changes from free to busy.

As shown in Table 1, detection ratio of proposed adaptive IDS is higher than traditional IDS against DoS attacks. Also, traditional IDS have a problem of False-Negative.

4.2 Scenario 2 Server State Changes from Busy to Free

In order to measure detection ratio between traditional IDS and proposed adaptive IDS, simulated Dos attack is conducted. Scenario 2 is used same environment like Scenario 1, simulation is conducted when state of server changes from busy to free.

Table 2 show result of simulation in state of server changes from busy to free.

As shown in Table 2, detection ratio of proposed adaptive IDS is higher than traditional IDS against DoS attacks. On the contrast of scenario 1, The False-Positive problem is occurred in scenario 2.

Table 2. Result of simulation in Server state: Busy -> Free

	Traditional IDS	Adaptive IDS
True-Positive	18	24
True-Negative	50	76
False-Positive	32	0
False-Negative	0	0

5 Conclusion

This paper suggested the IDS connected with MAPE model in order to implement the Environment Adaptive IDS. In the suggested methods used MAPE model to monitor any environmental changes and if any changes are spotted, environmentally adapted rule is generated. The environmentally adapted rule was applied to IDS. Since the rule generating method that the suggested methods used use pre-defined rule information, there exists limits to depend on the pre-defined rules. Due to such limits, the performance of Environment Adaptive IDS could be degraded. To overcome the limitations, researches about more flexible automatically generating rules are needed and if it is

achieved, it is highly expected that we would be able to operate the system which has the higher detection performance.

Acknowledgments. This research was supported by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT and future Planning(No. NRF-2016H1D5A1909989).

References

1. Pwc, Insight from The Global State of Information Security Survey. http://www.pwc.com
2. Roesch, M.: Snort: lightweight intrusion detection for networks. In: LISA, vol. 99, no. 1, pp. 229–238 (1999)
3. Wagner, D., Soto, P.: Mimicry attacks on host-based intrusion detection systems. In: Proceedings of the 9th ACM Conference on Computer and Communications Security, pp. 255–264. ACM (2002)
4. Kaur, S., Singh, M.: Automatic attack signature generation systems: a review. IEEE Secur. Priv. **11**(6), 54–61 (2013)
5. Mohammed, M.M., Chan, H.A., Ventura, N.: Honeycyber: automated signature generation for zero-day polymorphic worms. In: MILCOM 2008-2008 IEEE Military Communications Conference, pp. 1–6. IEEE (2008)
6. Li, Z., Sanghi, M., Chen, Y., Kao, M. Y., Chavez, B.: Hamsa: fast signature generation for zero-day polymorphic worms with provable attack resilience. In: 2006 IEEE Symposium on Security and Privacy, pp. 15–47. IEEE (2006)
7. Depren, O., Topallar, M., Anarim, E., Ciliz, M.K.: An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. Expert Syst. Appl. **29**(4), 713–722 (2005)
8. Peddabachigari, S., Abraham, A., Grosan, C., Thomas, J.: Modeling intrusion detection system using hybrid intelligent systems. J. Netw. Comput. Appl. **30**(1), 114–132 (2007)
9. De Lemos, R., et al.: Software engineering for self-adaptive systems: a second research roadmap. In: de Lemos, R., Giese, H., Müller, H.A., Shaw, M. (eds.) Self-Adaptive Systems. LNCS, vol. 7475, pp. 1–32. Springer, Heidelberg (2013). doi:10.1007/978-3-642-35813-5_1

OFART: OpenFlow-Switch Adaptive Random Testing

Dong-Su Koo[1] and Young B. Park[2(✉)]

[1] Department of Computer Engineering,
Dankook University, Juk-Jeon, Republic of Korea
ninedongsu@gmail.com
[2] Department of Computer Science,
Dankook University, Juk-Jeon, Republic of Korea
ybpark@dankook.ac.kr

Abstract. In the advent of SDN paradigm, the accumulated verification technologies in the existing software fields are being used to verify the SDN. Data Plane consists of Forwarding Devices and is controlled by Control Plane. If correctness of the Forwarding Device is not verified, it affects to the whole network. However, doing every testing by manually is a huge time-cost consuming act, so it requires an automation. In this paper, it suggests a framework which applies ART (Adaptive Random Testing) technique which considers OpenFlow Switch to be Black Box from the Controller point of view and is easy to do a testing automation.

Keywords: SDN · OpenFlow · Adaptive random testing

1 Introduction

SDN paradigm has brought us new challenges. SDN is a networking paradigm which deals with networking direction settings and controls or complicated operating management through software programming. SDN has acquired programmability through its operating ways and structure in which Control Plane and Data Plane are logically separated [1]. The Control Plane controls over Data Plane which consists of Data Forwarding Devices (switch, router, etc.) like Operating System. There is one which represents SDN: OpenFlow and it virtually is located in de-facto standard. The behavior of OpenFlow is that OpenFlow Control Plane, called OpenFlow Controller, updates Flow Table Entries in Data Plane, called OpenFlow Switch, and OpenFlow Switch is processed by flow based on the Flow Table [2]. So, since the role of software has grown huge, we could manage as well as control the network more flexibly and closely. However, the side effects of the software highly made the risk of having more defects. One of the network factors that could not be found from the beginning can affect to the whole network. According to severity of the defects, it would cause a serious result which ranges from packet loss to paralysis of the whole network. For example, if a switch actually run by network could not do a packet processing function, it would have caused a serial accident in the end. Thus, testing a correctness of network function is a very basically and fundamentally important activity.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_97

The testing or SDN-OpenFlow verification research which has been done above are substituting the testing methods or the existing software fields from the researchers' point of views towards SDN. However, most of the researches make SDN approach to Finite State Machine and only do Formal Verification by using authentication tools as in theorem prover, model checking [3–9]; it is actually difficult for testers to implement because of high learning curve. In the paper, we suggest functional testing methods in which OpenFlow Switch regards to Black-Box in a Controller point of view. Controller generates packets for a functional testing and put the packets in a Switch. Getting the results from the switch which contains the packets, compare them to ideal results. A way to generating packet used Adaptive Random Testing technique. In order to lessen time consuming, tedious functional testing characteristics, we suggest OpenFlow Switch testing automation framework which can automatize such as test execution, test generation, test selection, and test result comparison.

2 Background

2.1 Black-Box Testing

Black-box Testing is one of the Software Testing techniques that test software activities about internal software without any knowledge. The black-box testing is widely being used in almost every software [10]. The paper watches Switch as Black-box from the controller's point of view because when Ingress Packet is sent from the controller, black box switch plays a role to do as a packet and throw them back to Egress Packet.

2.2 Adaptive Random Testing

Random Testing can be applied to test generation, since it has the most popular and easiest advantages. However, for the Random Testing, as it tests randomly and selects, it has some performative problems in balancing between error detection and test generation. To fix these issues, Adaptive Random Testing is suggested by Cheng. ART generates test data more effectively and evenly based on Failure Pattern in Input Domain, and there is a variety of ways existed and is being researched. This paper uses FSCS-ART which chooses next test by distance in between a fixed size test candidate set and executed test [11, 12].

2.3 Testing SDN-OpenFlow Switch

SDN-OpenFlow verification and testing research are substituting existing software verification or testing methods from the researcher's point of view towards SDN. OFTEN is regarded as black-box, so it proceeds integrated network testing and detects unrevealed inconsistencies through state-space exploration techniques [13]. Through simple processes, we suggested approaches about integrated network testing, but by using the Model Checking technique, overhead factors about additional preliminary things are remained. SOFT is checking about Functional Equivalence of different

OpenFlow agents to test inter-compatibility. Individually different OpenFlow agents' functional equivalence is checking over whether they return the same results or not through symbolic execution [3].

Automated Test Packet Generation (ATPG) automatizes test packet generation based on topology information of organization and suggest the ways to minimize the number of test packets [14]. For FlowTest, by using Al Planning Tool, it generates Test Plan and suggests the way of testing Data Plane [15]. In case of InSP, it gets information from Packet Template Table and Flow Table and suggests the way of generating packet [16]. J. Yao et al. [4] provided a way and a tool that are able to test Black-Box in SDN Data Plane. Based on OpenFlow switch specification which supports multi table, formal model switch in a specification language, and after it makes Data Graph and Data Paths based on this model, it uses a way to creating test case through those two things. However, there are limits that the flow table status of an actual switch that constantly changes cannot be instantly applied to the switch model. Besides, for OFTest, it suggests a test case about OpenFlow Switch [17].

3 OFART: OpenFlow Switch Adaptive Random Testing

OFART consists of three big parts (1) Test Packet Generation, (2) Test Selection, (3) Test Execution. The flow chart is illustrated in Switch Testing Tool provided by SDN Open Source Framework RYU [18].

3.1 Overall

The whole OFART Framework architecture is illustrated in Fig. 1. At first, test generator creates the number *N* of Test Data and stores them into Test DB. Test Selector randomly chooses one of the Test Data out of *N*. After the Test Data which is chosen by Test Selector is created as a Test Pattern, is sent to Tester. Test Pattern is a test script file which is written in JSON format. Tester makes Test Pattern a packet and gives the packet to the switch which is to be tested. After the packet given to the switch is matched with Flow Table within the switch, output packets taking actions accordingly. The egress packet decides Pass or Fail after comparing Desired Packet with Egress Packet through Test Executor. Decided results are sent to Test Selector. The Test Selector chooses 10 candidates from Test DB, makes Candidate Set, and has distance value from recently executed Data. If compared results are Pass, Test Selector chooses Test Data which is the farthest distance. If it is Fail, it stores the Data in Fail DB, chooses the closest Test Data. The chosen Test Data is sent back to Generator and made as a test pattern. It repeats the acts shown above.

3.2 Adaptive Random Test Packet Generation

In this section, we describe of how Test Generator makes Test Pattern to make Test packet. Test Pattern is JSON which is to test OpenFlow Switch in RYU. Test Pattern consists of Prerequisite, Ingress Packet, Egress Packet. Test Generator generates Test

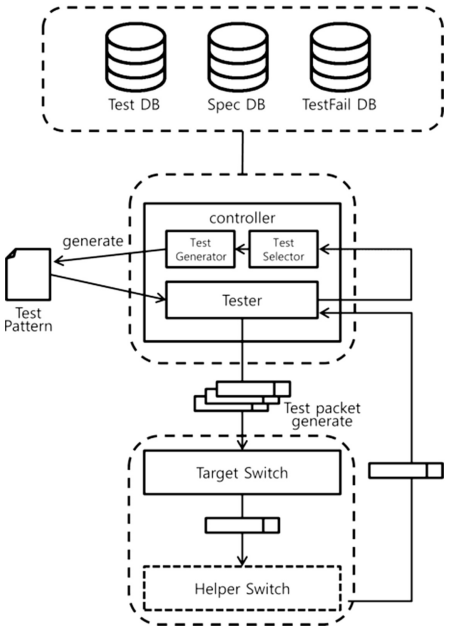


Fig. 1. High-level architecture of OFART framework

Pattern written in JSON format by ART technique. The created Test Pattern is sent to Tester and Tester interprets the Test Pattern and makes Test Packet to test switch. All the processes will be named as 'Adaptive Random Test Packet Generate', and acronym of the term is ARTPG.

3.3 Similarity-Based Test Selection

FSCS-ART picks up test data which will be executed next based on the distance between candidate set and executed test. The next data would be chosen as farthest data from the test data which was executed before. Numerical Input Domain can measure the distance through Euclidian distance, but for OFART, each factors of Test Data have a variety of shapes and scopes. In order to fix such problems, we defined a distance in between Tests based on similarity of tests. In other words, testing subset has a high similarity, and other functions a low similarity, Thus, tests with high similarity have a short weight: short distance, low similarity have a short weight: long distance.

Selector sends the test which is randomly chosen by Test DB to the Generator. And then, after calculating distances of the chosen test and randomly chosen 10 candidate tests, temporarily remember the farthest test, the closest test. After receiving a test result (Pass/Fail) through Generator and Test, if the result is Pass, sends the farthest test to Generator, or Fail, sends to the closest test to Generator.

4 Evaluation

In order to evaluate proposed method, we measure effectiveness using F-measure that is number of test cases required to detect the first defect. The defect is injected at a 'SET_DL_DST' function of Open vSwitch for F-measure. To compare traditional random testing with proposed method, the test case count and execution time are measured to find a injected defect. Setup time do not included in execution time. Table 1 shows simulation result of 100 times.

Table 1. Test result of random testing and OFART

	Random testing	OFART
Test case count	12	4
Execution time	0.924 s	0.346 s

As shown in Table 1, generated test cases of OFART are reduced 67 % than random testing. Also, execution time is reduced 62.5 %. As a result, OFART's performance show higher than random testing.

5 Conclusion

This paper applies ART technique in order for SDN Switch to do Functional Testing, suggests Framework which automatizes it. OFART can be used when Regression Testing and Tester are systematically verified in developed functions. Testers which will not involve in the development directly would be able to easily test the SDN Switch without implementing a high-learning curve. By applying ART technique, testing is under processing based on Failure Pattern, so it is more effective than the randomly pick-up way, and through automation, there would appear a reduction effect in the time and cost phases which are consumed by repeated actions of functional testing.

Acknowledgments. This research was supported by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (No. NRF-2016H1D5A1 909989)

References

1. Nunes, B.A.A., Mendonca, M., Nguyen, X.N., Obraczka, K., Turletti, T.: A survey of software-defined networking: past, present, and future of programmable networks. IEEE Commun. Surv. Tutorials **16**(3), 1617–1634 (2014)
2. Open Networking Foundation: Software-defined networking: the new norm for networks. ONF White Paper (2012)

3. Kuzniar, M., Peresini, P., Canini, M., Venzano, D., Kostic, D.: A SOFT way for openflow switch interoperability testing. In: 8th International Conference on Emerging Networking Experiments and Technologies, pp. 265–276. ACM (2012)
4. Yao, J., Wang, Z., Yin, X., Shiyz, X., Wu, J.: Formal modeling and systematic black-box testing of SDN data plane. In: 2014 IEEE 22nd International Conference on Network Protocols, pp. 179–190. IEEE (2014)
5. Guha, A., Reitblatt, M., Foster, N.: Machine-verified network controllers. In: ACM SIGPLAN Notices, vol. 48, no. 6, pp. 483–494. ACM (2013)
6. Skowyra, R., Lapets, A., Bestavros, A., Kfoury, A.: A verification platform for SDN-enabled applications. In: 2014 IEEE International Conference on Cloud Engineering (IC2E), pp. 337–342. IEEE (2014)
7. Ball, T., et al.: Vericon: towards verifying controller programs in software-defined networks. In: ACM SIGPLAN Notices, vol. 49, no. 6, pp. 282–293. ACM (2014)
8. Al-Shaer, E., Al-Haj, S.: FlowChecker: configuration analysis and verification of federated OpenFlow infrastructures. In: 3rd ACM Workshop on Assurable and Usable Security Configuration, pp. 37–44. ACM (2010)
9. Kang, M., Kang, E.Y., Hwang, D.Y., Kim, B.J., Nam, K.H., Shin, M.K., Choi, J.Y.: Formal modeling and verification of SDN-OpenFlow. In: 2013 IEEE Sixth International Conference on Software Testing, Verification and Validation, pp. 481–482. IEEE (2013)
10. Jorgensen, P.C.: Software Testing: a Craftsman's Approach. CRC Press, Hoboken (2016)
11. Chen, T.Y., Leung, H., Mak, I.K.: Adaptive random testing. In: Maher, M.J. (ed.) ASIAN 2004. LNCS, vol. 3321, pp. 320–329. Springer, Heidelberg (2004). doi:10.1007/978-3-540-30502-6_23
12. Chen, T.Y., Kuo, F.C., Merkel, R.G., Tse, T.H.: Adaptive random testing: the art of test case diversity. J. Syst. Softw. **83**(1), 60–66 (2010)
13. Kuzniar, M., Canini, M., Kostic, D.: OFTEN testing OpenFlow networks. In: 2012 European Workshop on Software Defined Networking, pp. 54–60. IEEE (2012)
14. Zeng, H., Kazemian, P., Varghese, G., McKeown, N.: Automatic test packet generation. In: 8th International Conference on Emerging Networking Experiments and Technologies, pp. 241–252. ACM (2012)
15. Fayaz, S.K., Sekar, V.: Testing stateful and dynamic data planes with FlowTest. In: Third Workshop on Hot Topics in Software Defined Networking, pp. 79–84. ACM (2014)
16. Bifulco, R., Boite, J., Bouet, M., Schneider, F.: Improving SDN with InSPired Switches (2016)
17. OFTest. http://www.projectfloodlight.org/oftest
18. RYU. https://osrg.github.io/ryu

An Evaluation of Availability, Reliability and Power Consumption for a SDN Infrastructure Using Stochastic Reward Net

Kihong Han[1], Tuan Anh Nguyen[1,2], Dugki Min[1(✉)], and Eun Mi Choi[3]

[1] School of Computer Engineering, Konkuk University, Seoul, South Korea
gksrlghdl@gmail.com, dkmin21@gmail.com, anhnt2407@konkuk.ac.kr
[2] Office of Research, University-Industry Cooperation Foundation, Konkuk University, Seoul, South Korea
[3] School of Management Information Systems, Kookmin University, Seoul, South Korea
emchoi@kookmin.ac.kr

Abstract. Networking infrastructure of a software defined network (SDN) is demanding further studies to achieve continuity and high availability of data transactions for cloud computing services. Nevertheless, the high-speed and complicated network of hosts and network devices often encounters with a variety of failures either of links or system components. This paper aims to study the specific characteristics of and the impact of various failures on a typical SDN infrastructure. We propose a stochastic model using stochastic reward nets (SRN) with the incorporation of hardware failures (of hosts, switches, storages and links) and software failures (virtual machines (VM). The system model is analyzed based on steady state availability in the case of default parameters. Comprehensive sensitivity analyses are conducted to study the system behaviors with regard to different major impacting factors. Reliability analysis is also carried out to pinpoint the role of VM migration in extending the system lifetime. System power consumption based on availability of every module is also conducted to examine the power allocation on system devices and operations. This study provides a helpful basis for network design and implementation in SDN infrastructure.

Keywords: Software defined network · Stochastic reward net · Availability · Reliability · Power consumption

1 Introduction

Software defined network (SDN) has emerged as a promising networking paradigm to resolve the existing limitations of the current network infrastructure in modern cloud computing systems. The core idea of SDN is to separate the network control logic (control plane) from the physical networking devices (routers, switches) which perform data forwarding functionalities (data plane) [1, 2]. Thus, the network infrastructure is

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_98

simplified to consist of simple forwarding devices whereas network control logic is implemented in a network controller (or network operating system). This concept highlights the programmable capability and flexible reconfiguration of the SDN to simplify policy enforcement. In fact, physically centralized network infrastructure still requires adequate levels of system availability and reliability.

Although many efforts were made to enhance the availability and reliability of the SDN, the previous work are mostly focused on the implementation of the SDN in different manners. The core technology and applications of virtualization for SDN are of focus in various studies. For instance, the server virtualization and virtualized systems as the fundamental elements of SDN are studied in detail in [3]. A network of virtualized servers complying a data center network topology is studied in [4] and the use of virtualized systems for disaster tolerance is introduced in [5]. On the other perspective, the scalability of a network virtualization in the SDN are analyzed in [6], and different data forwarding algorithms for protection and restoration to improve the network resiliency against cyber-attacks and other component failures were proposed in [7]. Also, many others tackled various security issues of the SDN including the pros and cons of the SDN security [7], and designing a secure and dependable SDN control platform [8]. The work [9] dealt with network management challenges, as well as failures and recovery of single/multiple controllers [10–12]. A few work assessed SDN performance through experiments [13]. The dependability of SDN with respect to the number of network devices and hosts was presented in [14] with more focus on SDN controller rather than SDN physical infrastructure. Generally, previous works do not provide a common ground to assess the availability and reliability of SDN. Therefore, our effort in this paper is to conduct a comprehensive evaluation of the SDN infrastructure.

The main contributions of this paper are summarized as follows:

- study detailed characteristics and system behaviors of networking infrastructure of a typical SDN. We incorporate a variety of failure modes including server failures, virtual machines (VM) failures, and switch failures. To avoid any severe loss, we incorporate the VM live migration between clusters of servers or between direct servers.
- propose availability model using stochastic reward nets (SRN). The availability model is modified to create a reliability model for the sake of system reliability assessment. The model is also used in power consumption assessment.
- comprehensively assess the network infrastructure of the SDN based on various metrics of interest including steady-state availability (SSA), sensitivity of system availability, system reliability and power consumption over time of major subsystems and operations.

The rest of this paper is organized as follows. Section 2 introduces a typical architecture of the SDN infrastructure in consideration. Section 3 presents the proposed SRN model of the SDN. Section 4 presents the numerical results and system assessment. Section 5 highlights some existing issues to discuss and for future work. Section 5 concludes the paper.

2 A SDN Architecture

Figure 1 depicts a typical SDN architecture in consideration. The system consists of control plane and data plane as common design concept of SDNs [4]. Currently, the control plane does not involve in the modelling and analysis in this paper. The data plane (as also known as SDN infrastructure) composes of a certain number of physical devices including SDN networking devices (switches S1, S2, and S3), 6 physical machines (also called hosts: (H00, H01), (H10, H11), (H20, H21)). The networking in the SDN infrastructure is implemented either for SDN management or for fault-tolerance. Particularly, the physical machines are organized in clusters, in which every two machines connect to a switch (H00 and H01 connect to S0; H10 and H11 connect to S1; H20 and H21 connect to S2). Furthermore, physical machines are equipped a multiple-port network card to either connect to switch in a cluster and to connect directly to another machine in other clusters (H00 connects to H10; H11 connects to H20; H21 connects to H01). The network bandwidths of the internal and external connections within and between clusters are different. Connections within a cluster (between hosts and a switch of the cluster) are much faster than those between clusters (between hosts of different clusters). The SDN controller communicates with the switches via a typical protocol (e.g. Open Flow [15–17]) for data packet forwarding functionalities. Whereas, system users can access the machines through switches. The physical machines are virtualized to host a number of VMs for computing services. Users might run apps on the VMs. We will model and analyse this system architecture using analytical modelling in the next sections.

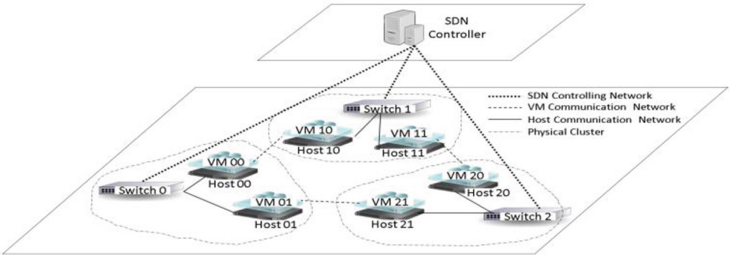


Fig. 1. A typical SDN architecture

3 Stochastic Reward Net Model

This section presents in detail the stochastic modeling of the system using SRN.

3.1 System SRN Model

The whole system SRN model is shown as in Fig. 2. The model consists of several submodels for hardware and software components. The physical devices (switches and hosts) are modeled using two-state models (up and down state models). The software

subsystem (VMs) are modeled together in a single submodel to capture the interaction and communication between the VMs of the SDN. A certain state of a component is represented by a blank circle. If a number presents in the circle that means the component currently stays in the corresponding state. In the VM subsystem model, if there is a multiple number in a certain circle, it means a number of VMs all stays in the same corresponding state. The blank rectangle represents for the timed-transition in which the transition between states consumes a period of time (e.g. time to failure, time to repair, time to migration etc.). Whereas the black-shaded rectangle represents for the immediate transition in the way that a transition between states is considered as instant occurrence regardless of time. This is to capture the dependent behaviors of the VM subsystem on the physical components (e.g. when a host enters downstate, all of its hosted VMs are instantly migrated to mitigate downtime services). The system initiates with all subsystems in upstate. In order to capture the exact behaviors of the system, a set of guard functions is employed in order to enable/disable transitions and manipulate the transitions of tokens. The complex combination of token locations in the SRN model forms the corresponding system state. Thus every transition of any token from a place to another place in the system model implies the corresponding transition of system states.

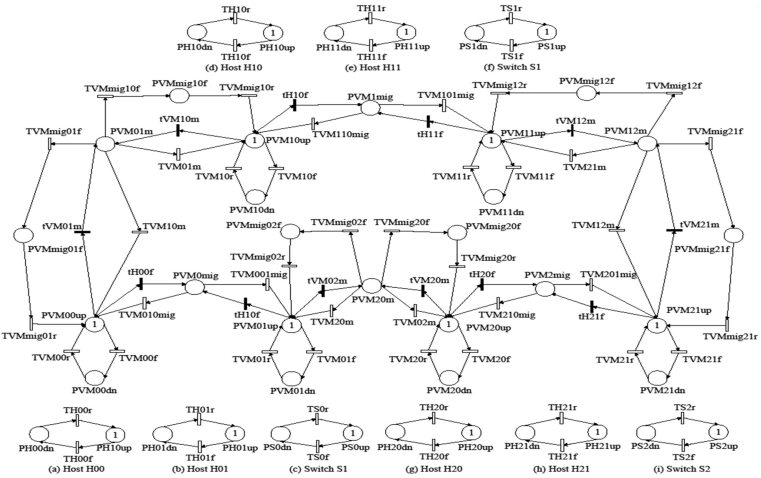
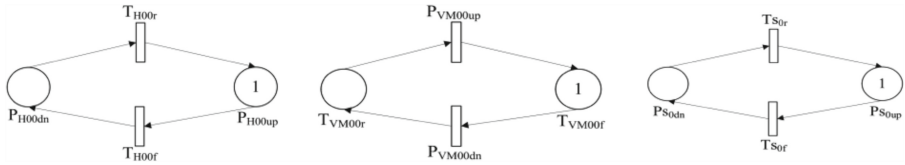


Fig. 2. SRN system model

3.2 Two-States SRN Models of Hosts and VMs

Figure 3 shows the two-states models of the non-redundant subsystems with only one kind of outage/recovery in the cluster 0 including host H00 (Fig. 3a), virtual machine VM00 (Fig. 3b) and switch S0 (Fig. 3c). The other hosts, VMs and switches in the remaining clusters are modeled identically. From the normal states (P_{H00up}, P_{VM00up}, P_{S0up}), the models go to down state upon corresponding hardware or software failures with the rates respectively λ_h, λ_{vm}, λ_s. Accordingly, a token is taken out from the aforementioned upstate and deposited in the corresponding downstates

(P_{H00dn}, P_{VM00dn}, and P_{S0dn}) through the red timed-transitions T_{H00f}, T_{VM00f} and T_{S0f}. The models return upstate after the completion of respective recovery with the rates μh, μ_{vm}, and μs. As soon as the timed-transition represented for recovery actions T_{H00r}, T_{VM00r} and T_{S0r} are enabled, the tokens in the downstates are removed and deposited back in the respective upstates. The corresponding subsystem comes back its operational state. In the submodel of VM subsystem on a certain host (Fig. 3b), there could be a multiple number of VMs at the same time in the upstate P_{VM00up}. In this case, these VMs compete to each other to fail in advance thus with higher rate dependent on the number of existing VMs in the upstate P_{VM00up} at a certain time. This behavior is captured using marking dependence in the way that the rate of the timed transition T_{VM00f} varies upon the number of tokens in the place P_{VM00up}. The marking dependence is denoted by the mark '#' besides the timed transition T_{VM00f}.

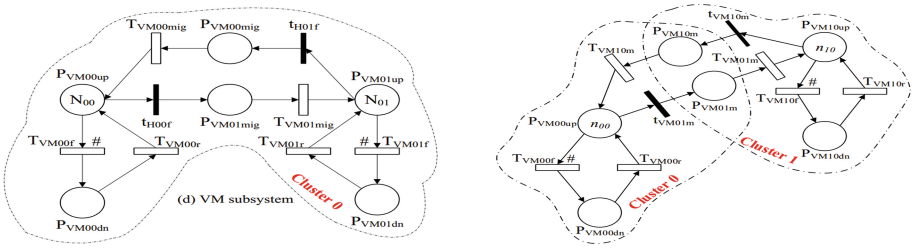


(a) Host H0 (b) VM00 (c) Switch S0

Fig. 3. Two-states SRN models of host, VMs and switches

3.3 VM Live-Migration SRN Models

i. Within a Cluster: Figure 4a shows the modeling of VM live migration within a certain cluster. The migration is triggered when one of the host goes down. Particularly, we consider the live migration between the hosts H00 and H01 in the cluster 0. When the host H00 goes down (one token is deposited in the place P_{H00dn} in Fig. 3a) and the host H01 operates in healthy state, the immediate transition t_{H00f} res to remove all tokens in the place P_{VM00up} and deposit them in the intermediate place $P_{VM01mig}$. This is to say that all the running VMs suddenly go down, the live migration process is triggered instantly and the VM image files on the network memory system are subsequently migrated onto the healthy host with the rate ω_{mig}. When the migration process of a VM completes, the timed-transition $T_{VM01mig}$ is enabled and one token is taken out from the place $P_{VM01mig}$ and deposited in the place P_{VM01up}. The migrated VM enters its operational state but hosted on the remaining host. The similar VM live migration process goes identically when the host H01 goes down and the host H00 is in upstate. The immediate transition t_{H01} fires, then all tokens in the place P_{VM01up} are taken out and deposited in the place $P_{VM00mig}$. The migration process with the rate ω_{mig} deposits a token at a time into the place P_{VM00up}. The VM of the failed host H01 is migrated completely onto the healthy host H00. The intermediate places $P_{VM00mig}$ and $P_{VM01mig}$ represent for the network memory system within a cluster where the synced VM image files are stored for fault-tolerance. The within-cluster VM live migration is applied for the pairs of hosts in clusters (H00, H01); (H10, H11) and (H20, H21) as



(a) Within a Cluster (b) Between Clusters

Fig. 4. SRN models of VM live-migration

shown in the system model (Fig. 2). The migration is conducted via high-speed network connections (about Gb/s) within a cluster.

ii. Between Clusters: The between-cluster VM migration is modeled as in Fig. 4b. The migration process is triggered instantly in two cases: (i) a switch fails and (ii) both hosts of a cluster stay in downtime. Either any of these cases happens, the system users are not able to access their VMs. Therefore, it is necessary to migrate the users' VMs onto the other cluster to improve the system availability to the users. The migration depicted in Fig. 4b is between the host H00 of the cluster 0 and the host H10 of the cluster 1. In particular, when one of the aforementioned failures occurs in the cluster 0 supposedly, the immediate transition t_{VM01m} is enabled. All tokens in the place P_{VM00up} are removed and deposited in the place P_{VM01m}. The between- clusters migration is triggered consequently. When the migration completes, the timed-transition T_{VM01m} is enabled to remove one token from the place P_{VM01m} and deposited in the place P_{VM10up}. The VM on the inaccessible cluster is now migrated on the healthy host in the accessible cluster. If the failure cases occur on the cluster 1, the VM migration is similarly implemented and captured by the immediate transition t_{VM10m}, the place P_{VM10m} and the timed-transition T_{VM10m}. The identical rates (ω_m) of the timed-transitions T_{VM01m} and T_{VM10m} represent for the network speed of the connections between the clusters. With the assumption that, the clusters locates in different places in a cloud data center, the value of this network bandwidth is likely smaller than that of the network bandwidth of the local connections within a cluster.

4 Numerical Analyses and Result

The SRN system model is implemented in the Stochastic Petri Net Package (SPNP) [18]. Default input parameters are summarized as in Table 1. Our metrics of interest include steady state availability, availability sensitivity, system reliability and power consumption of the system.

Table 1. Default input parameters for SRN model

Name	Attached transition	Description	Meantime/Values
$1/\lambda_H$	T_{Hf}	Mean time to failure of a host	1 month
$1/\mu_H$	T_{Hr}	Mean time to repair a host	8 h
$1/\lambda_{VM}$	T_{VMf}	Mean time to failure of a VM	7 days
$1/\mu_{VM}$	T_{VMr}	Mean time to repair a VM	1 h
$1/\lambda_S$	T_{Sf}	Mean time to failure of a switch	6 months
$1/\mu_S$	T_{Sr}	Mean time to repair a switch	24 h
ω_{mig}	T_{VMmig}	Network bandwidth of the connections within a cluster	1 Gb/s
ω_m	T_{VMm}	Network bandwidth of the connections between clusters	256 Mb/s
S_{VM}		Image file size of a VM	10 GB
C_h		Power consumption of a host per time unit	500 W
C_{vm}		Power consumption of a VM per time unit	76 W
C_{sw}		Power consumption of a switch per time unit	223 W
C_{mig}		Power consumption of VM migration per time unit	25 W

4.1 Availability Analyses

(a) Steady State Availability Analyses: To compute the steady state availability (SSA), a reward is assigned to every marking (i) of the SRN model. The reward r_i is defined as in the Eq. 1:

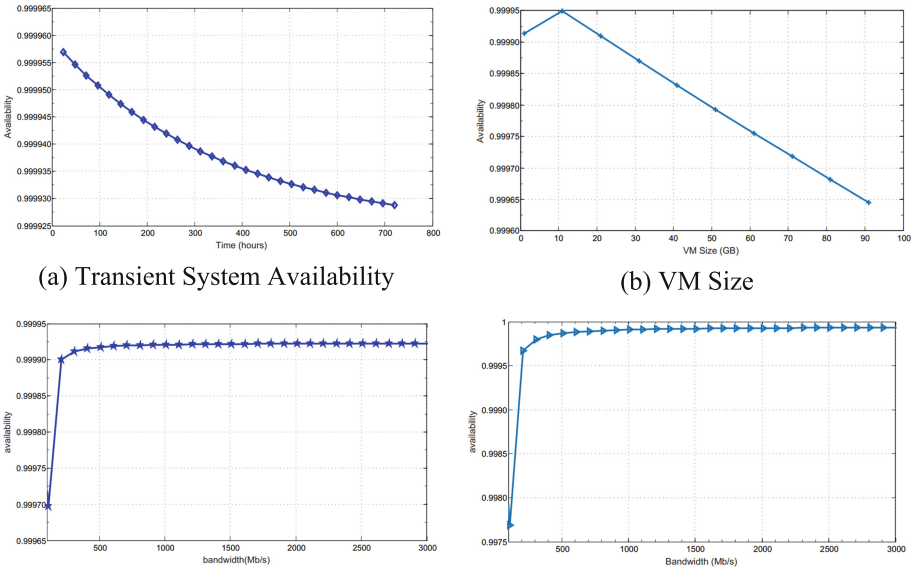
$$r_i = \begin{cases} 1 & \text{if } (\#P_{VM00up} + \#P_{VM01up} + \#P_{VM10up} + \#P_{VM11up} + \#P_{VM20up} + \#P_{VM21up} > 0) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

We compute the SSA using SRN for two cases: (i) the SDN with one isolated cluster (SDN01) and (ii) the proposed SDN (SDN02). We limit the number of the VMs in the above cases at 2 VMs, since our focus is to highlight the capability of the SDN to rescue the VMs onto the other hosts for business continuity in use in any unexpected failures of physical hardware.

The output result of SRN method show that the overall system availability is improved vastly when the system configuration changes from one cluster (SDN01) to three cluster (SDN02) even though the number of VMs are same in both configurations. This is to say that networking in the SDN infrastructure is important not only for data forwarding functionalities but also for fault-tolerance. Since, if a failure occurs in any severe manner (both hosts go down or switch fails in one cluster), the VMs in services are likely rescued to other clusters with little downtime of the service.

(b) Availability Sensitivity Analyses: The sensitivity analyses are conducted by varying a specific input parameter while fixing all the other parameters. This type of analysis provides proper assessment on the system characteristics over the variation of parameters. Thus, it helps find the bottle-neck of the input parameters. Our parameters in interest are the operational time of the whole system, network bandwidth (within and between clusters) and the size of VM image. The results are as follows.

- *System time*: Fig. 5a shows the variation of system's overall availability over time. As time goes by, the system availability gradually decreases. The value goes down faster in early period of time (0-300] hours when many failures of different system components probably occurs. In the late time, the availability approaches a steady value which is actually the steady state availability as computed in Table 2.
- *VM size*: The dependence of the system availability upon the VM image le size is shown clearly as in Fig. 5b. Under the default values of the network bandwidths, the size of VM image le at about 10 GB (as default value in Table 1) is the optimal to obtain the highest system availability. If we intend to migrate the VMs with either smaller or bigger size, the system availability actually decreases. With smaller VM size than the optimal value, even though the migration time could be mitigated but the migrated VMs depend on the new hosts' operational states. If the VMs have bigger size, it likely takes longer time for migration and thus it is not efficient to gain higher availability for the VMs.
- *Network bandwidths*: Fig. 5c and d present the variation of the system availability respectively upon the changes of network bandwidths ω_m and ω_{mig}. Basically, both



(a) Transient System Availability

(b) VM Size

(c) ω_m

(d) ω_{mig}

Fig. 5. Availability sensitivity analyses of the SDN infrastructure

Table 2. Steady state analysis of the SDN

Case	SRN
SDN01	0.997161819972
SDN02	0.999913646747

types of network bandwidths (within and between clusters) bear the same effect on the system availability. In particular, in the early stage (0–500] Mb/s, an increase of the bandwidths can vastly boost the system availability. Whereas in the late period (over 500 Mb/s), the above-mentioned effects do not maintain. The system availability stays stable regardless of the increase or decrease of the network bandwidths. This can be explained as follows. *(i)* Under the default value of the VM size (10 GB), the system needs to find the optimal values of the bandwidths in order to migrate the VMs but also to gain highest availability. Thus, a small increase of the network bandwidths in early stage can actually provoke a huge increase of system availability. *(ii)* As soon as the system finds the optimal values of the bandwidths, the even faster network connections do not contribute more effectiveness of the VM migration processes in general. Furthermore, it is noticed that the variations of the network bandwidth between clusters (ω_m) can enhance higher system availability in comparison with the variations of the network bandwidth within a cluster (ω_{mig}). These system characteristics are the significant base in designing the SDN infrastructure.

4.2 Reliability Analysis

In order to evaluate the system reliability over time, the availability model in Fig. 2 is modified to create a reliability model in the way that all recovery actions for hosts, VMs and switches are removed. The analysis result is shown in Fig. 6. Since the system is assumed to not have any recovery actions, the system reliability obviously falls down along with the passing of time. The whole system goes down at about 1000 h after its start. This also pinpoints that the VM live-migration techniques employed in the system do extend the lifetime of the system beyond the lifetime of the physical hosts ($1/\lambda_H$ = 720 h).

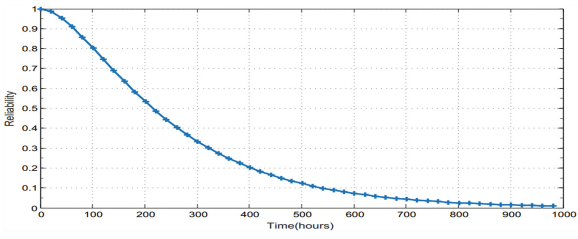


Fig. 6. A typical SDN architecture

Table 3. Power consumption in a year

Module/Operation	Power (Watt)
Hosts	2,997,393.29
VMs	151,624.96
Switches	669,660.57
Migration	3.55
Overall	3,818,682.38

4.3 Power Consumption

The operations of the system actually consume a certain amount of energy in Watt (W) over time. Without loss of generality, we assume that the major sources of power consumption are on the system modules (hosts, VMs, switches) and main system operations (VM migration) while all the other energy leaks are not main our focus and ignored consequently. Thus, the overall power consumption of the system is defined as in the Eq. 2:

$$P_{SDN} = P_h + P_{vm} + P_{sw} + P_{mig} \tag{2}$$

$$P_h = C_h.A_h.T \tag{3}$$

$$P_{vm} = C_{vm}.A_{vm}.T \tag{4}$$

$$P_{sw} = C_{sw}.A_{sw}.T \tag{5}$$

$$P_{mig} = C_{mig}.A_{mig}.T \tag{6}$$

Where:

- C_h, C_{vm}, C_{sw}, C_{mig} are respectively the power consumption in a time unit (W/h).
- A_h, A_{vm}, A_{sw}, A_{mig} are respectively the overall availability of the modules or operations over a period of time.
- T is a certain period of time in consideration.

We will use the Eqs. 2–6 to compute the overall system's power consumption over time. The constant values are listed in Table 1 referenced from [19, 20]. Table 3 shows the power consumption of every system components and the overall SDN infrastructure in consideration over a year. As the physical devices, hosts and switches consume the most power. Since we run VMs on the system along the way in a year, it also consumes some amount of power. The overall power consumption in a year of the SDN infrastructure is about 3.8 Mega Watt (MW).

5 Conclusion

This paper has shown a comprehensive modeling and analysis of a typical SDN infrastructure using SRN. The system was assessed based on a variety of metrics of interest including steady-state availability, availability sensitivity, reliability and system power consumption. The results pinpointed that to obtain the high-availability infrastructure for a SDN, it is necessary for the system to comply with a specific network topology and to apply different fault-tolerant methods (e.g. VM live migration).

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the University Information Technology Research Center support program (IITP-2016-R2720-16-0004) supervised by the IITP (Institute for Information & communications Technology Promotion)

References

1. Hu, F., Hao, Q., Bao, K.: A survey on software-defined network and OpenFlow: from concept to implementation. IEEE Commun. Surv. Tutorials **16**(4), 2181–2206 (2014). doi:10.1109/COMST.2014.2326417
2. Cui, Y., Xiao, S., Liao, C., Stojmenovic, I., Li, M.: Data centers as software defined networks: traffic redundancy elimination with wireless cards at routers. IEEE J. Sel. Areas Commun. **31**(12), 2658–2672 (2013). doi:10.1109/JSAC.2013.131207
3. Nguyen, T.A., Kim, D.S., Park, J.S.: A comprehensive availability modeling and analysis of a virtualized servers system using stochastic reward nets. Sci. World J. 1–18 (2014). doi:10.1155/2014/165316
4. Nguyen, T.A., Min, D., Park, J.S.: A comprehensive sensitivity analysis of a data center network with server virtualization for business continuity. Math. Probl. Eng. (2015). http://www.hindawi.com/journals/mpe/aa/521289/
5. Nguyen, T.A., Kim, D.S., Park, J.S.: Availability modeling and analysis of a data center for disaster tolerance. Future Gener. Comput. Syst. **56**, 27–50 (2016). doi:10.1016/j.future.2015.08.017
6. Drutskoy, D., Keller, E., Rexford, J.: Scalable network virtualization in software-defined networks. IEEE Internet Comput. **17**(2), 20–27 (2013). doi:10.1109/MIC.2012.144
7. Vaghani, R., Lung, C.-H.: A comparison of data forwarding schemes for network resiliency in software defined networking. Procedia Comput. Sci. **34**, 680–685 (2014). doi:10.1016/j.procs.2014.07.097
8. Kreutz, D., Ramos, F.M., Verissimo, P.: Towards secure and dependable software-defined networks. In: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking - HotSDN 2013, p. 55 (2013). doi:10.1145/2491185.2491199
9. Kuklinski, S.S., Chemouil, P.: Network management challenges in software-defined networks. IEICE Trans. Commun. **E97-B**(1), 2–9 (2014)
10. Li, H., Li, P., Guo, S., Nayak, A.: Byzantine-resilient secure software-defined networks with multiple controllers in cloud. IEEE Trans. Cloud Comput. **2**(4), 436–447 (2014). doi:10.1109/TCC.2014.2355227
11. van Adrichem, N.L., van Asten, B.J., Kuipers, F.A.: Fast recovery in software-defined networks. In: 2014 Third European Workshop on Software Defined Networks, pp. 61–66. IEEE (2014). doi:10.1109/EWSDN.2014.13
12. Yao, G., Bi, J., Guo, L.: On the cascading failures of multi-controllers in software defined networks. In: 2013 21st IEEE International Conference on Network Protocols (ICNP), pp. 1–2. IEEE (2013). doi:10.1109/ICNP.2013.6733624
13. Gelberger, A., Yemini, N., Giladi, R.: Performance analysis of software-defined networking (SDN). In: 2013 IEEE 21st International Symposium on Modelling, Analysis and Simulation of Computer and Telecommunication Systems, pp. 389–393. IEEE (2013). doi:10.1109/MASCOTS.2013.58
14. Longo, F., Distefano, S., Bruneo, D., Scarpa, M.: Dependability modeling of software defined networking. Comput. Netw. **83**, 280–296 (2015). doi:10.1016/j.comnet.2015.03.018
15. Xia, W., Wen, Y., Foh, C., Niyato, D., Xie, H.: A survey on software-defined networking. IEEE Commun. Surv. Tutorials **PP**(99), 1 (2014). doi:10.1109/COMST.2014.2330903
16. Xie, J., Guo, D., Hu, Z., Qu, T., Lv, P.: Control plane of software defined networks: a survey. Comput. Commun. **67**, 1–10 (2015). doi:10.1016/j.comcom.2015.06.004
17. Farhady, H., Lee, H., Nakao, A.: Software-defined networking: a survey. Comput. Netw. **81**, 79–95 (2015). doi:10.1016/j.comnet.2015.02.014

18. Ciardo, G., Muppala, J., Trivedi, K.: SPNP: stochastic Petri net package. In: Proceedings of the Third International Workshop on Petri Nets and Performance Models, PNPM 1989, pp. 142–151 (1989). doi:10.1109/PNPM.1989.68548
19. Marcu, M., Tudor, D.: Power consumption measurements of virtual machines. In: 2011 6th IEEE International Symposium on Applied Computational Intelligence and Informatics (SACI), pp. 445–449. IEEE (2011). doi:10.1109/SACI.2011.5873044
20. Huang, Q., Gao, F., Wang, R., Qi, Z.: Power consumption of virtual machine live migration in clouds. In: 2011 Third International Conference on Communications and Mobile Computing, pp. 122–125. IEEE (2011). doi:10.1109/CMC.2011.62

A Dynamic Service Binding Scheme with Service OID for IoT

Euihyun Jung(✉)

Department of Computer Science, Anyang University, 602-14, Jungang-ro Buleun-myeon, Ganghwa-gun, Incheon, Korea
jung@anyang.ac.kr

Abstract. Object Identifier (OID) has been used in various areas, but its usage has been limited to naming things for a long time. In order to extend OID for identifying Web services, we suggested a new kind of OID named Service OID and its relevant resolution scheme. It is expected for devices to easily collaborate by resolving a simple number-shaped Service OID to the corresponding URI. This simple resolution seems straightforward, but it is not enough for a caller to use the converted URI because every URI needs its unique parameter signature. For this, in the proposed resolution scheme, a data schema was designed and used to provide the binding information of URI.

Keywords: Object Identifier (OID) · Internet of Things (IoT) · Dynamic service binding

1 Introduction

Object Identifier (OID) is a globally unique ISO (Internationalized Organization for Standardization) identifier [1, 2]. It has a reverse tree namespace that enables organizations to make their own unique namespaces under the allocated tree arc to the organizations in an independent manner [3]. OID has served to name things in diverse areas such as Health Management [4], Network Management [5], X.509 certificates [6], etc. OID has useful features and it is specialized in naming things, but researchers have not paid much attention to extend OID to identifying virtual things.

If OID is extended for this purpose, the convergence of physical and virtual things can be much easier and this will result into the rise of new application areas. Actually, URI is de facto standard for devices to call each other's services, but URI occasionally tends to be long and complex. Comparing to URI, OID has an advantage of its simplicity. With only simple number-shaped OID similar to a phone number, users can indicate a virtual service without knowing the complex URI of the service. This feature will also help the collaboration of devices in Internet of Things (IoT) [7], because the devices can call each other's services with only OID numbers.

However, in order to realize this vision, several prerequisites have to be resolved. First of all, an OID should be translated into a URI indicating the target service. For this, an identifier resolution system has to be provided. Secondly, the binding information for this converted URI should be given after the resolution process of the OID has been completed. The first task seems not difficult because the existing OID

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_99

Resolution System (ORS) can be easily upgraded to add this kind of the resolution. Comparing to the first one, the second task will be a headache because callers have to understand the parameters and the signature of the URI before using the service connected to the URI. In order to tackle this issue, the resolution of OID should contain a scheme that gives callers binding information of the target URI. The task seems not easy to solve, but if it is done, OID-based service binding for IoT will be feasible.

In this paper, we proposed a new kind of OID named *Service OID* that can indicate any kind of Web services with simple numbers and dots. To resolve Service OID, we designed a scheme that provides a protocol and a data schema for service binding between peers having OID only. The rest of this paper is organized as follows. Section 2 describes a proposed architecture, a binding schema and the operation of the system. Section 3 concludes the paper.

2 System Design

2.1 Service OID and Resolution Scheme

OID has widely used in identifying any kind of things. It is based on the hierarchical name structure of the OID tree [2]. This OID naming uses a sequence of names delimited by dots while traversing the OID tree from the root. In the root level, there are 3 nodes; itu-t(0), iso(1), and joint-iso-itu-t(1). Each node of the OID tree has both a name and a number and it can have the infinite level of sub nodes under itself. For example, the OID of "wmo" is {joint-iso-itu-t(2).alerting(49).wmo(0)}. It represents the "joint-iso-itu-t" node has an "alerting" sub node and the "alerting" node has a sub node of "wmo". Also, OID can be represented as a sequence of numbers delimited by dots from the root node. In this example, the OID of "wmo" will be "2.49.0".

While the main purpose of OID is indicating things, Service OID is proposed to identify Web services. Using Service OID, a caller can use easily service providers' Web services with simple number-shaped OID. To make it possible, some kind of resolution scheme is needed. As shown in Fig. 1, we proposed the resolution scheme which has three functional parts; A Root ORS, a Service ORS and a service provider. When a caller wants to use a service indicated as a OID like "1.4.3.2", the caller requests a resolution to the Root ORS with the OID. Then, the Root ORS resolves the OID with the corresponding URI of the Service ORS. The caller queries the Service ORS how the binding information of the URI consists. After getting the binding information from the Service ORS, the caller requests the service provider for the target service.

2.2 Service Binding

URI Signature. When a caller wants to use a Web service with its URI, the caller has to know how to assign the arguments of the URI in advance. For example, a service which provides the geographical information of the current location will demand its callers to assign proper location data to the corresponding parameters. However, practically, the

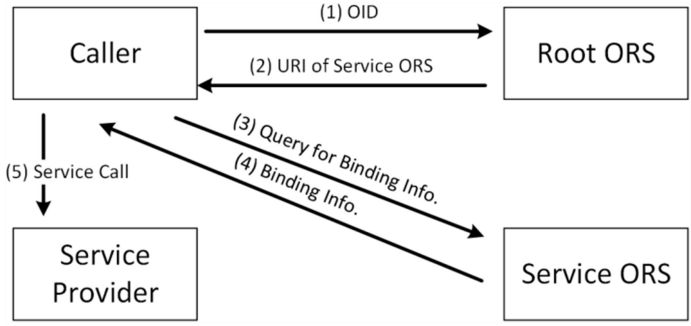


Fig. 1. The proposed scheme consists of three parts which collaborate to resolve a service OID.

callers cannot figure out how many parameters the URI has and which data type each parameter follows. In the current technology, the only solution is embedding the hardcoded binding codes in the caller to use the service. Definitely, it will result in limiting dynamic binding and mash-up services between devices. In this paper, to resolve the issue, the Service ORS was designed to give binding information which describes the parameters and the signature of the URI.

Design of a schema. To describe the signature of the URI, an XML schema was designed. The schema consists of two parts. The first part is the signature of a target URI. The part defines a method, a return type, a URI template [8], and a description. The second part describes the name and the data type of the parameters written in the URI template. Table 1 shows each part of the schema.

A Sample Service Binding. Binding information with the proposed schema tells a caller how URI template consists and how parameters are assigned. For example, a service for the MD5 (Message Digest 5) hashing can have a sample binding information in shown Fig. 2. Let's assume the service has an OID of "1.4.3.2". In the scenario, a caller that wants to use the MD5 service will call the service with the OID. Then, the root ORS returns the Service ORS's URI such as "http://service.example.org/md5". The caller requests the service ORS to give the binding information for the

Table 1. The schema for the description of the URI signature

Attribute	method[a]	GET or POST		
	return_type[a]	MIME type for the returned data		
	template[a]	URI template		
	desc[b]	Human readable text describing the URI		
Child node	parameter[c]	Attribute	name[a]	The parameter name included in the signature
			data_type[a]	The data type of the parameter
			desc[b]	Human readable text describing the parameter

[a]mandatory, [b]0 or 1, [c]more than 0 (appeared in binding information).

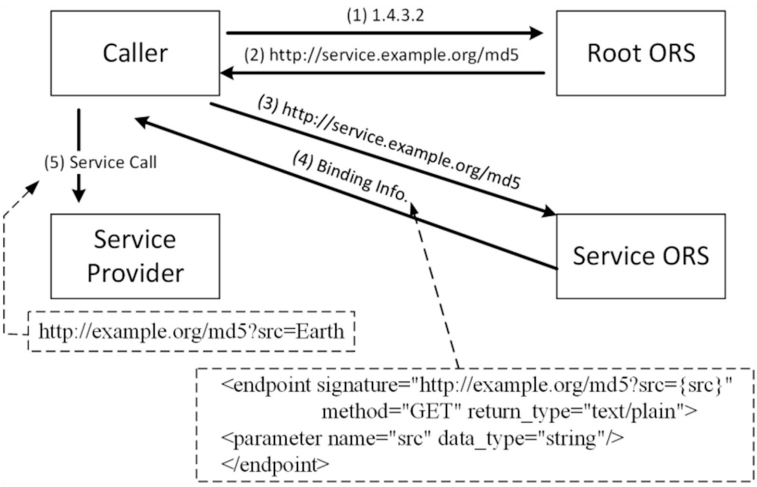


Fig. 2. The sample MD5 resolution shows how a Service OID is resolved and the target service is called.

service with the given Service ORS's URI. After getting the binding information from the Service ORS, the caller finds parameters from the signature part of the information. In this case, the parameter is "src" and its type is described in the <parameter> node. Then, the caller assigns the parameter "src" with "Earth" string and calls the service such as "http://example.org/md5?src=Earch".

A Sample Application. A sample application has been made to show the feasibility of the proposed scheme. In the sample application, the mobile app shows its users the locations of near hospitals and pharmacies. As shown in Fig. 3, the mobile app just only knows the OID of the service instead of the exact service URI. When its users

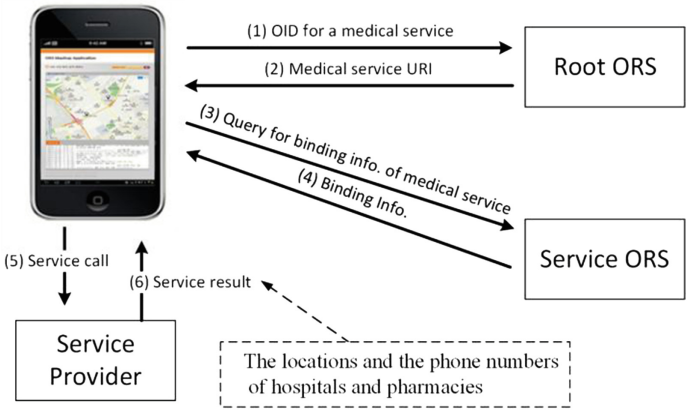


Fig. 3. The sample application uses the Service OID indicating to the medical service to get the detailed information of the medical institutes.

initiate the mobile app, it just sends the OID to the resolution infrastructure, which resolves the OID to the target service. Then, the mobile app calls the service to display the locations and other information of medical institutes.

3 Conclusion

Although OID has been used in various areas, its usage has been limited to naming things for a long time. In this paper, we suggested Service OID and its relevant resolution scheme to extend OID. Service OID is a new kind of OID to identify Web services. It is expected for devices in IoT environment to collaborate with each other using a simple number-shaped Service OID instead of a complex URI. However, the simple resolution from Service OID to URI cannot provide a caller with binding information. In order to give the binding information to the caller of Service OID, the resolution scheme consisting of the protocol and the data schema was designed. In the future, semantic-based binding information will be researched for the intelligent services of IoT.

References

1. Dubuisson, O.: Introduction to Object Identifiers (OID) and Registration Authorities. http://www.oid-info.com/doc/introductiontoobjectidentifiers(OIDs).pdf. France Telecom Orange
2. ITU-T committee: Information technology – Procedures for the operation of object identifier registration authorities: General procedures and top arcs of the international object identifier tree. ITU-T X.660 (2011)
3. ITU-T committee: Open systems interconnection – Object identifier resolution system (ORS). ITU-T X.672 (2010)
4. Steindel, S.J.: OIDs: how can I express you? Let me count the ways. J. Am. Med. Inf. Assoc. **17**(2), 144–147 (2010)
5. Lee, D.Y., Kim, D.S., Pang, K.H., Kim, H.S., Chung, T.M.: A design of scalable SNMP agent for managing heterogeneous security systems. In: Network Operations and Management Symposium, pp. 983–984 (2000)
6. Myers, M., Ankney, R., Malpani, A., Galperin, S., Adams, C.: X. 509 Internet public key infrastructure online certificate status protocol-OCSP. IETF RFC-2560 (1999)
7. Bandyopadhyay, D., Sen, J.: Internet of Things: applications and challenges in technology and standardization. IEEE Wirel. Pers. Commun. **58**(1), 49–69 (2011)
8. Gregorio, J., Fielding, R., Hadley, M., Nottigham, M., Orchard, D.: URI Template. IETF-RFC 6570 (2012)

A Context-Aware Architecture Pattern to Enhance the Flexibility of Software Artifacts Reuse

Doohwan Kim, Soon-Kyeom Kim, Woosung Jung, and Jang-Eui Hong(✉)

School of Electrical and Computer Engineering, Chungbuk National University, Chungdaero 1, Cheongju 28644, Republic of Korea
{dhkim, skkim}@selab.cbnu.ac.kr,
{wsjung, jehong}@chungbuk.ac.kr

Abstract. In software development approaches including open source-based development, new version development of software product, reuse has become a useful and common approach. However, the problem of traditional reuse approach is not easy to find reusable components that software engineers want, or difficult to reuse the components without any modification. This paper proposes an ICFP (Imparter-Collector-Fetcher-Presenter) architectural pattern that supports context-aware reuse in order to solve the existing problems in the reuse process for the software artifacts. The proposing reuse approach based on the ICFP pattern provides the benefits of improving the flexibility of reusing methods as well as enhancing precision of component retrieval.

Keywords: Artifact reuse · Context-aware · Architecture pattern · Flexible reuse

1 Introduction

Reuse is an activity to use existing and proven assets in some forms within the software product development process [1]. Recently, the reuse-based software development has recognized as a major and important approach to develop software products because of increased open source-based development and shorten product lifecycle [2, 3]. However, There are some problems as like that the users (software engineers) could not easy to find a suitable component which is satisfying the users' desired or not easy to reuse the component without any modification even a reusable component was found [4, 5]. In order to solve these problems, a variety of methods and concepts have been proposed to reuse the components, and such techniques of software architecture [6], software product line [7] are also proposed to maximize the reusability of existing assets. However, these problems occurring from the process of software reuse are still remained.

This research was supported by NRF Korea funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030505).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_100

This paper focus on the reuse of the text-based software artifacts, which have their own contents, as defined in the DID (Data Item Descriptions) of MIL-STD-498 [8]. These documents developed in a project are managed as a file with the whole contents, and also the entire document is accessed as a reusable unit. Therefore, software engineers have to look through the whole document to decide whether the document is enough to reuse or not. But this process is not efficient to support reuse in easy and systematic way because the software engineers have to scroll up and down the document to reuse, and also a candidate document may contains irrelevant or various information.

Therefore, this paper proposes a reuse approach based on ICFP (Imparter-Collector-Fetcher-Presenter) pattern in order to resolve the problems of the existing reuse approaches. Especially, the ICFP pattern is developed with the concept of context-aware reuse, and also defined with the extension of the existing MVC (Model-View-Controller) architecture pattern [9]. The proposed pattern supports software engineers to find more suitable component which they want, and supports to recommend the reusable candidates which are more designated toward reuse subject as the results of document retrieval.

Our proposed pattern can provide the benefits of improving easy to use, precision, and flexibility for software engineers in the process of reuse-based software development. The ICFP pattern will be utilized to build a repository system which can support efficient and flexible reuse in a software project, which is a R&D (Research and Development) project of software system including vast amount of literature, technical documents, articles, and software artifacts.

2 Reusable Component and Framework

Our study defines a context-aware reuse framework to support fine-grained reuse of software artifacts, as shown in Fig. 1.

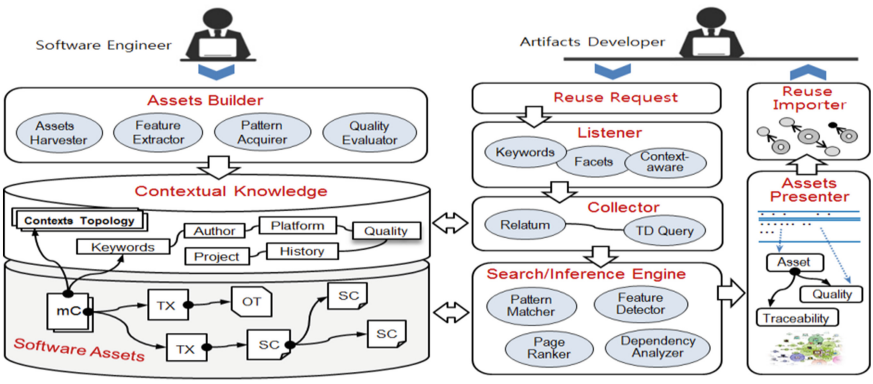


Fig. 1. The context-aware reuse framework which represents major components and their relationships of reuse repository.

As shown in Fig. 1, the 'Asset Builder' has a function to register and extract the reusable components from software documents or source codes, the 'Listener' is a function to process the reuse requests from the users, the 'Search Engine' searches components from reusable asset storage according to user's request. And the 'Presenter' represents the search results in predefined display styles. In addition, the components registered for reuse are stored in the physical storage, and the contextual information for these components is configured separately in the repository. Among these constituents, we briefly describe the properties [10] of reusable components.

Reusable component as defined in this study is a section (or content) unit of a document or a unit function in source codes rather than entire document or source file. We named this reusable component as microComponent (mC). The characteristics of the mC can be summarized as follows;

- mC is stored and managed by a section unit extracted from a document.
- mC is expressed with XML syntax for its structure.
- mC has the relationships between the previous and the next mCs within a document.
- mC has also the relationships between the preceding and succeeding documents.
- mC has the constraints that should be considered when reuse.
- mC has the history counter to indicate the number of reuse.

Construction of component repository based on the mC provides the following advantages: (1) It is possible to retrieve more suitable components to the users' desire. (2) The size of retrieval result (i.e., the number of the reusable candidates) is relatively small, thus the users can fast determine whether to reuse. And (3) the size of reuse unit can be extended in flexible manner.

3 Contexts-Aware Architectural Pattern

In order to support easy and flexible reuse, we define a context-aware architecture pattern, called as ICFP pattern. This pattern is designed with an extension of existing MVC pattern, as shown in Fig. 2. As shown in Fig. 2, the ICFP pattern is composed of Reuse Imparter, Collector, Fetcher, and Presenter. Their brief functionalities were explained in below.

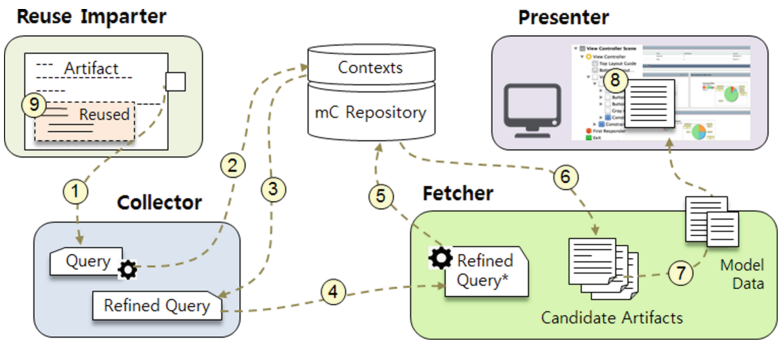


Fig. 2. The ICFP architectural pattern (dynamic view) to support flexible reuse

Reuse Imparter: When a user requests the reuse in the process of creating a software artifact, the 'Reuse Imparter (RI)' identifies the cursor position in the current creating document, and then gathers the XML tag information. The acquired information from XML tags is a combination of the document title, the chapter title, and the section titles, which provides the contextual information of currently creating document. RI sends (①) the acquired information to the 'Collector'.

Relatum Collector: The 'Collector' selects (②) first the relevant relata from ontology base using the contextual information which is maintained the repository, and then generates (③) a final query for the requested reuse from the user. The final query is sent (④) to the 'Fetcher'.

Component Fetcher: The 'Fetcher' selects (⑤) the appropriate mCs from the repository using the query sent by the Collector. And it creates (⑥) a list of reusable candidates to show the list to the user. The 'Fetcher' will push (⑦) the list to the 'Presenter' depending on the 'Model Data' which is a meta-data for the detail information of mC, required to visualize the candidate components.

Result Presenter: The 'Presenter' displays (⑧) the information for the candidate components using a visualization method that is dependent on the type of mC. The visualization method is determined with the context information of the mC. Based on this information, the user can import (⑨) the selected component to reuse into his/her creating artifact.

4 Flexible Reuse

The flexible reuse, which is proposed in this paper, means that the users can dynamically decide the granularity (i.e., the size) of a component to reuse mCs. In order to support this type of reuse, (1) it must be provided with the relationship information between mCs to provide the traceability among them. (2) It must be also defined a mechanism that dynamically extends the granularity of an mC.

4.1 Traceability Between Reusable Components

In order to support flexible reuse, it needs the traceability between mCs, which is provided with the connection information of the previous and the next mCs, and the preceding and succeeding mCs. Figure 3 depicts the logical connections between mCs. That is, the variables 'Previous_mC' and 'Next_mC' are used for tracing mCs within a document, and the variables 'Preceding_mC' and 'Succeeding_mC' are used for tracing mCs between two documents.

4.2 Dynamic Extension of Reusable Component

After the reusable candidates were chosen by context-aware retrieval, the user selects a component (i.e., mC in this paper) from the candidates, and then determines whether

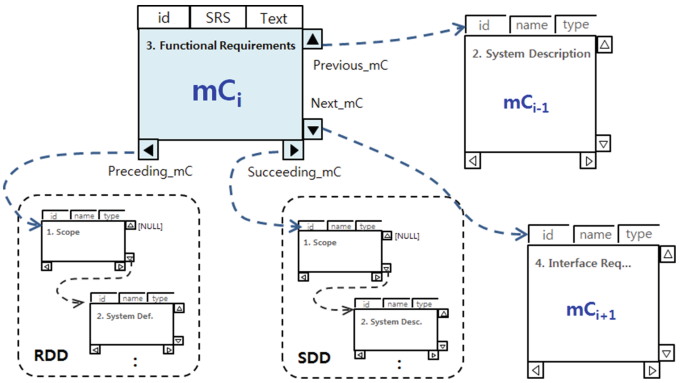


Fig. 3. The logical connections between mCs to provide the traceability

the component will be reused or not through checking the detail information of the component. In our general case, an mC does not exceed to 1 page or at most 10 pages because an mC is corresponding to a section of a document. So, the user can easily determine whether it is reused or not for a candidate component.

When trying to reuse a particular mC, the user can examine the previous mC and the next mC to reuse together because those mCs are closely relate to the particular mC. This situation can be supported with our flexible reuse approach. Figure 4 shows the type of patterns for extending the reuse granularity.

The brief usage of each pattern, as shown in Fig. 4, is as followings.

- Self: in case of that the selected candidate mC is complied with user's desire in the topic, thus user tries to reuse of the mC only.
- Extend-Up: In case of that user tries to extend the reuse range to the previous section.
- Extend-Down: In case of that user tries to extend the range to the next section.
- Extend-Both: In case of that user tries to extend the range to both directions with the previous mC and the next mC.
- Extend-Full: In case of that user attempts to reuse the entire document that contains the current selected mC.

A user can extend the reuse granularity of mC based on these patterns, and insert the extended mC into the cursor position of currently creating artifact. The history counter will be increased, and a pop-up window will be appeared to record the user's review, at the time point of the insertion.

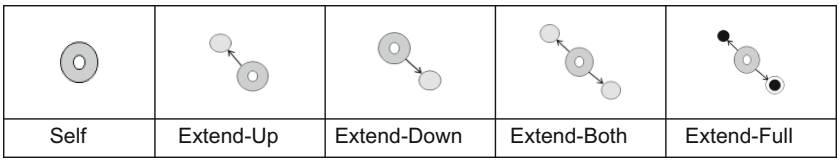


Fig. 4. The patterns to extend the granularity of reusable component

5 Conclusion and Further Work

We defined a reusable component as a microComponent (mC) that is a section unit of a document or a function unit of source codes rather than a whole document or a source file, respectively. This mC is to support easy, fast and flexible reuse of software artifacts. Based on the mC concept, this paper proposed ICFP architecture pattern to support context-aware reuse of software artifacts.

Our proposing reuse approach was intended to develop a context-aware reuse framework and to build-up a repository for software artifact reuse. Based on our framework, we can provide more narrow results of component retrieval, which is more suitable to users' desire. This makes it possible to fast decision for whether a selecting mC will reuse or not. Also in order to provide flexible reuse of the mC, we also defined the extension patterns of the granularity for an mC. The results of this study are to improve the precision for reusable component search, after all to increase the reusability of software artifacts.

Our follow-up work of this study is to develop a technique of learning-based component fetch from component repository. We believe that this fetch technique will be feasible with the learning for the users' behavioral patterns; the history information and the used extension pattern for a particular mC will be used for more intelligent retrieval and selection in the reusable component repository.

References

1. Weide, B.W., Ogden, W.F., Zweben, S.H.: Advance in Computers: Reusable Software Components. Academic Press, London (1991)
2. Heinemann, L., et al.: On the extent and nature of software reuse in open source Java projects. In: The 12th International Conference on Software Reuse (ICSR 2011), pp. 207–222 (2011)
3. Frakes, W.B., Kang, K.: Software reuse research: status and future. IEEE Trans. Softw. Eng. **31**(7), 529–536 (2005)
4. Agresti, W.W.: Software reuse: developers' experiences and perceptions. J. Softw. Eng. Appl. **4**(1), 48–58 (2011)
5. Schmidt, D.C.: Why Software Reuse has Failed and How to Make It Work for You. C++ Rep. Mag., January 1999
6. Gomaa, H., Farrukh, G.A.: Composition of software architectures from reusable architecture patterns. In: The 3rd International Workshop on Software Architecture, pp. 45–48 (1998)
7. Gery, E., Scouler, J.L.: Strategic reuse and product line engineering. IBM developerWorks, pp. 1–17 (2014)
8. MIL-STD-498: Data Item Descriptions (DIDs) for MIL-STD-498, Department of Defense, USA, January 1995
9. Plakalovic, D., Simic, D.: Applying MVC and PAC patterns in mobile applications. J. Comput. **2**(1), 65–72 (2010)
10. Kim, D.H., Kim, S.K., Hong, J.-E.: A microComponent-based reuse technique for reusing software architecture. In: The 11th Asia Pacific International Conference on Information Science and Technology, pp. 209–212, June 2016

Deep Analysis of Tag Interference by Tag to Tag Relative Angles with Passive Far Field UHF RFID System

Jae Sung Choi and Hyun Lee(✉)

Department of Computer Engineering, Sunmoon University, Sunmoon-ro-221, Tangjeong-myeon, Asan-si, ChungCheongnam-Do 31460, Korea
{jschoi, mahyun91}@sunmoon.ac.kr

Abstract. This paper study about Passive UHF RFID tag to tag interference with consideration of tags' relative angle and distance. According to the analysis results of the measurement, the RSSI values under tag interference are significantly affected by tags' posture angles. And this analysis will be used for tag interference model developments.

Keywords: Tag interference · Passive UHF RFID · RSSI · Relative angle

1 Introduction

In Location Base Service (LBS) system, Real Time Localization System (RTLS) for object is a critical and essential issue to provide maximized tracking and visibility feasibilities. Although a homogeneous passive UHF far-field RFID systems with GEN 2 protocol [1] are widely used in supply chain and logistic applications, the passive RFID could not be a primary solution of location sensing of target of interest due to lack of accuracy of localization. In this paper, we study one of the reason of inaccurate and unprecise performance of the passive far-field UHF RFID based localization system.

Cost efficiency is the biggest advantage of the passive RFID based localization system compared to other type of active device based localization systems. However, because the passive RFID system uses backscattering communication, the communication range between tag and the reader's antenna is short. Moreover, accuracy of the localization depends on environmental factors, and the tag to tag interference is one of most significant factor. In [2], the authors evaluated the decrease of interrogation range by multiple adjacent tags' interference. In [3], the adjacent tag affects to decrease of reader's interrogation range. This issue is also related to the tag detuning problem when the tag adheres on a metallic object.

According to our previous work [4], since a passive RFID tag approaches a certain existing tag (as a reference tag), the existing tag's IC impedance is affected by the approached tag, and the tag's backscattered signal strength is significantly changed. The reason of the effect is a metal tag antenna impacts to an adjacent tag's backscattered signal strength, when the tag places in certain distance. Figure 1 shows the degree of interference at RSSI of existing tag. In this paper we study more about the tag to tag interference with consideration of tag to tag relative angles and various integration time duration.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_101

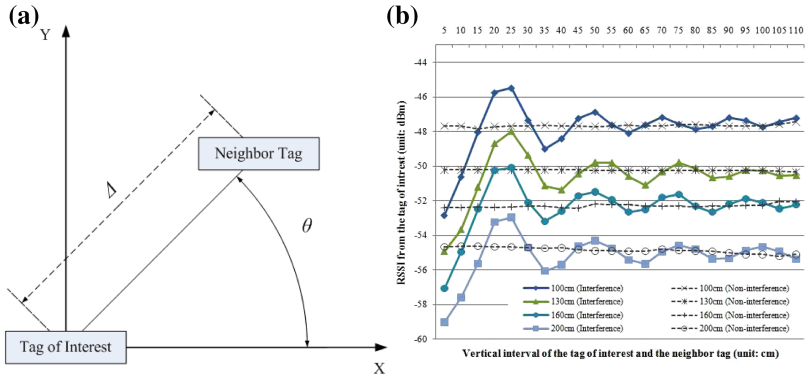


Fig. 1. Interferences of adjacent tag with tag to tag distances (Δ) and angles (θ) in an anechoic chamber, where $10\ \text{cm} \leq \Delta \leq 75\ \text{cm}$, and $0° \leq \theta \leq 90°$. RSSI fluctuations under tag-to-tag-interference, where adjacent tag locates vertically from TOI [4].

The rest of this paper is organized as follows. In Sect. 2, we present a system set-up for the measurement of tag interference with various operating frequency and geometric relations. In Sect. 3, we analyze the tag interference and we discuss our conclusions and future work in Sect. 4.

2 System Set-up

For analysis of tag-to-tag interference, we set-up the test environment in an anechoic chamber as shown in Table 1. Tag interference had been analyzed with the relation such as relative distance and angle in 2-D conditions. Tag of Interest (TOI) is fixed at a position with 200 cm between the reader antenna and the tag. Figure 2(a) illustrates two orientations as θ_1 and θ_2 for the tags.

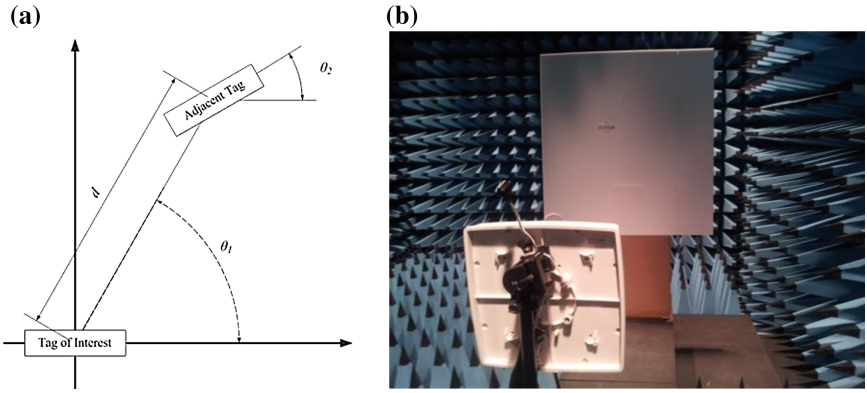


Fig. 2. Test set-ups for analysis of tag interferences. (a) Impact of the relative angles (orientation) between two tags. (b) Test set-up in an anechoic chamber

Table 1. Specification of system setup.

Specification	Details
Reader	SIRIT INFINITY 510
Antenna	Poynting Patch-A0025 with 6 dBi of gain
Target Tags	Alien Squiggle Gen 2 Passive tags
Frequency	915 MHz
TX Power	30 dBm (1 W)
Antenna Location	80 cm above the ground
Antenna to Tag distance	200 cm
Environment	In an anechoic Chamber

3 Analysis Results

We analyze the impact of the relative angles (orientation) between two tags to RSSI. Figure 2(a) illustrates two orientations as θ_1 and θ_2 for the tags. Depending on the angles, degree of interference can be varied and it relates to RSSI variation at certain physical locations for the tag as shown in Figs. 3 and 4. Figure 3(a) to (d) denote tag interference when θ_1 is fixed and θ_2 is varied as 0 to 90°.

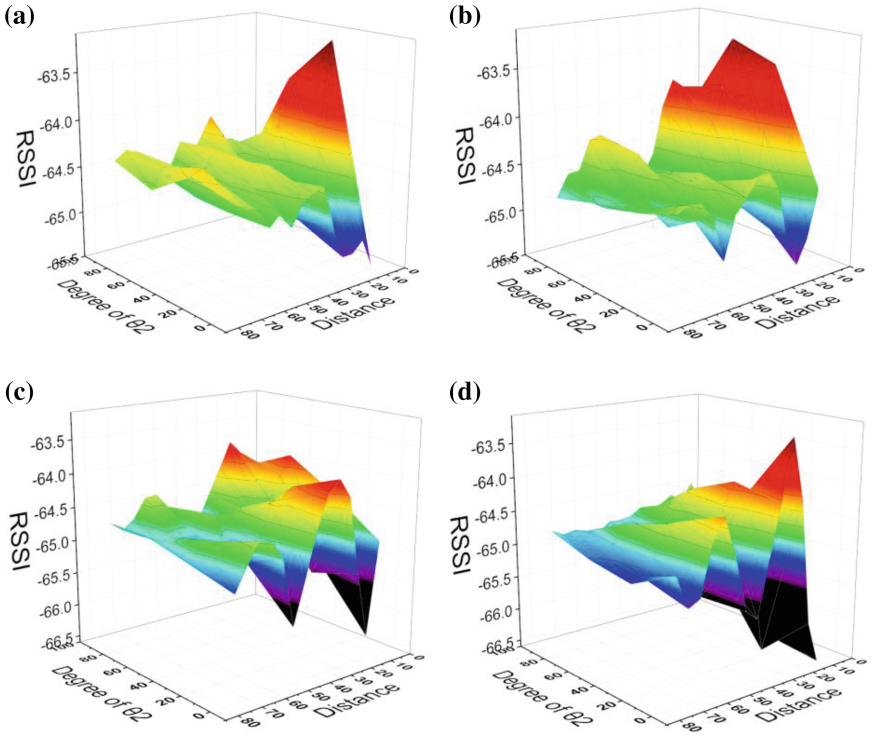


Fig. 3. RSSI measurements of Tag of Interest under fixed degree of θ_1. (Unit of RSSI: dBm, unit of tag to tag distance: cm) (a) $\theta_1 = 0$. (b) $\theta_1 = 30$. (c) $\theta_1 = 60$. (d) $\theta_1 = 90$.

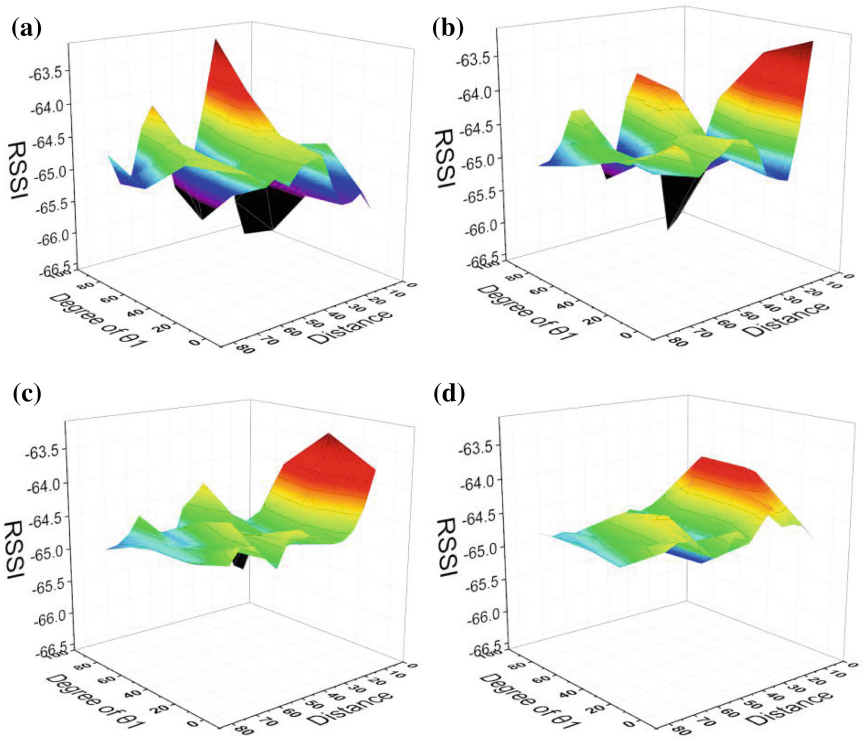


Fig. 4. RSSI measurements of Tag of Interest under fixed degree of θ_2. (Unit of RSSI: dBm, unit of tag to tag distance: cm) (a) $\theta_2 = 0$. (b) $\theta_2 = 30$. (c) $\theta_2 = 60$. (d) $\theta_2 = 90$.

As shown in Fig. 3(a), when the degree of θ_1 is fixed most of case, the fluctuations of interference of RSSI are shown as exponentially decreased sine wave form. In detail, when θ_1 is 0°, increase of θ_2 brings out more variation of interference. Another feature in Fig. 3 is a level of height in the fluctuation. When the θ_1 is set as 60° and 90°, the level of height is sharply increased. Especially, when the θ_1 is 90°, the level is most significantly changed. The maximum difference of RSSI is 3.401dB when θ_1 is 90°, and θ_2 is 0°.

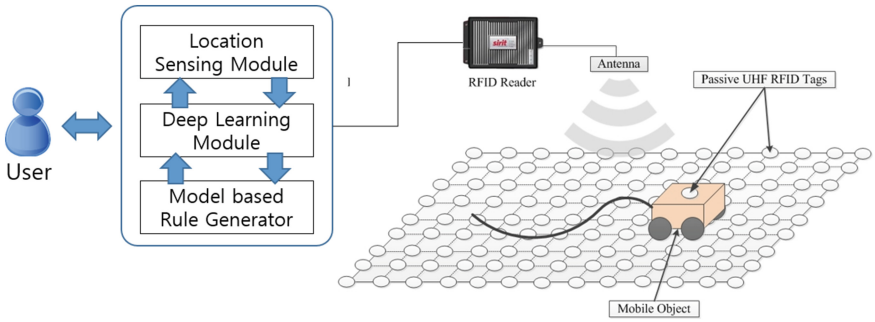


Fig. 5. Tag to Tag interference model based LBS system

Figure 4 illustrates fluctuations of interference of RSSI when the degree of θ_2 is fixed. On the contrary to fixed θ_1 cases, when the θ_2 is set as 0° and 30°, the level of height is sharply increased. Especially, when the θ_2 is 0°, the level is most significantly changed.

4 Conclusion

In passive UHF RFID systems, we presented of adjacent tag influences to another tag, and we defined the tag-to-tag interference problem. When two tags locate close each other, an adjacent tag influences other tags' IC impedance. And it causes excess increase or decrease of the backscattering communication budgets. According to our observation in our experimentation, tag-to-tag interference changes the reader received signal strength is changed over 3.4dB depended on the distance and relative angle between two tags in the anechoic chamber.

For the future work, the observed tag-to tag interferences between the two tags, as the target of interest, and the adjacent tag have to be modelled with their relative angles. As shown in Fig. 5, the models will be applied to Passive UHF RFID based real-time object localization system for supply chain application and inventory management applications.

Acknowledgments. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013R1A1A1075980).

References

1. GEN2. http://www.gs1.org/epcrfid/epc-rfid-uhf-air-interface-protocol/2-0-1
2. Dobkin, D., Weigand, S.: UHF RFID and tag antenna scattering. Part I: experimental results. Microwave J. **47**(5), 170–190 (2006)
3. Dobkin, D., Weigand, S.: UHF RFID and tag antenna scattering. Part II: theory. Microwave J. **49**(5), 86–96 (2006)
4. Choi, J.S., Kang, M., Elmasri, R., Engels, D.W.: Investigation of impact factors for various performances of passive UHF RFID system. In: 2011 IEEE International Conference on RFID-Technologies and Applications, pp. 152–159 (2011)

Simple Method of Video Mapping of Multiple Targets

In-Jae Jo[1], Joohun Lee[2], and Yoo-Joo Choi[1,3(✉)]

[1] Department of Newmedia, Seoul Media Institute of Technology, Seoul, South Korea
yjchoi@smit.ac.kr
[2] Department of Newmedia Contents, Dong-Ah Institute of Media and Arts, Anseong, South Korea
[3] Immersive Media Laboratory, Seoul Media Institute of Technology, Seoul, South Korea

Abstract. In this paper, a simple and efficient projection mapping framework is proposed so that even a projection mapping novice can easily map multiple video clips onto multiple different flat surfaces of the real world, which are not orthogonal to the projection direction. We also propose a symmetrical mesh refinement method to reduce the severe distortion of each image resulting from a flat surface having a certain angle to the projection direction. Through experiments, we proved that the proposed method reduced the image distortion error by 95.19 %.

Keywords: Projection mapping · Multiple video mapping · Texture distortion

1 Introduction

Projection mapping, known as video mapping and spatial augmented reality, is a technology of projecting graphic images onto the surface of an irregularly shaped object such as building, indoor object, or theatrical stage. This technology has been widely used in advertising, exhibition, media art, performance, etc., since it makes the visual content immersive by providing dynamic changes to the fixed real environment. Recently, interactive projection mapping has attracted much interest that various approaches have been actively researched [1, 2]. In performing arts in particular, fantastic stages have been produced interactively according to the motion of the dancers or actors by using real-time projection mapping technology. For interactive projection mapping, motion-sensing input devices such as Kinect should be installed, and calibration between motion-sensing space and 3D projection space has to be carried out before real-time projection mapping is performed. These procedures are not simple, and it is a great challenge to recognize exactly meaningful gestures of multiple persons in a large stage in computer vision.

In general projection mapping called spatial augmented reality, not an interactive one, there are still some hurdles that should be overcome by using technologies of computer graphics and image processing. These obstacles are attributed to the fact that the real environment should be controlled with 2D or 3D virtual objects together.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_102

Bimber, et al. explained that, in contrast to traditional virtual reality, augmented reality attempts to embed synthetic supplements into the real environment instead of immersing a person into a completely synthetic world [3]. They also pointed out that a real environment is much more difficult to control than a completely synthetic one. Therefore, in general, projection mapping content is controlled by using specialized software such as Madmapper, Millumin, and Resolume Arena to fit any desired image onto the surface of an object existing in the real environment. Note, however, that these software tools have some limitations for inexperienced users. First of all, relatively long training experience is required for general users to use these tools ably. Moreover, more complicated procedures are necessary to project multiple video clips onto several different surfaces of objects. As another important drawback, the projected graphical images even onto the flat surface of an object may be distorted. This happens because the projected graphical geometry unit, such as triangle or rectangle, should be deformed when the normal surface of the object is not orthogonal to the projection direction. Since such triangle or rectangle is matched with a texture, this deformation causes inevitable distortion of image. In numerous cases, the rectangle geometry is used for the mapping of a video clip, and a rectangle is generally rendered as two triangles in GPU (Graphic Processing Unit). Therefore, if the rectangle is largely deformed by the projection direction, the texture image extracted from a video clip and mapped onto the rectangle geometry can be seriously distorted on the border of the two triangles. In most specialized mapping tools, users define a refined geometric mesh in order to map a virtual image onto a 3D uneven surface or to reduce the distortion of the image mapped onto the flat surface. Users control the texture deformation by manually moving the mesh points. It is a very time-consuming and unnecessary task in case of mapping onto a flat surface since the deformed positions of internal mesh points can be automatically calculated by linear interpolation among boundary points.

In this paper, we present a projection mapping framework that allows even a projection mapping novice to map multiple video clips onto any flat surfaces of the real world easily. We also propose a method that allows an image to be mapped onto a flat surface without severe distortion in all projection directions. Through experiments, we proved that image distortion is reduced by applying the proposed method.

2 Proposed Method

2.1 Design of Video Mapping Classes for Multiple Targets

Projection mapping is sometimes called media façade in media art, which is made by combining “media” and “façade” that means the front wall of a building. Similarly, media façade uses an external wall of a building as canvas for drawing and displays multimedia content on it by projection-based or LED-based method. On the other hand, projection mapping exploits any kind of surfaces on which light can be projected, including external walls, interior spaces, and even isolated objects. Project mapping involves more than the projection-based display method, however. Though various surfaces are used as drawing canvas, only one media source is mostly projected onto the surfaces of the real environment at any one time. In order to project multiple video

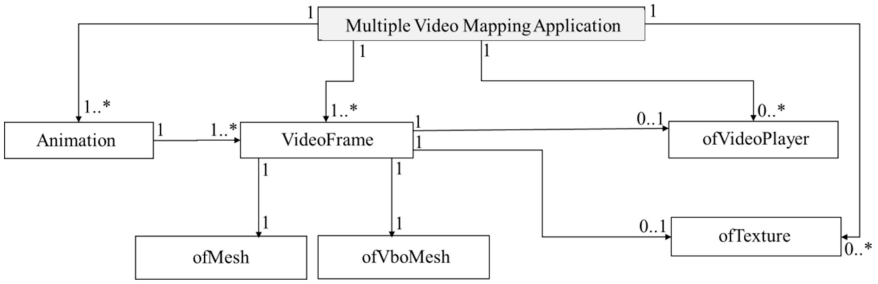


Fig. 1. Class diagram of the proposed system.

sources simultaneously onto different surfaces, more complicated procedures are needed in existing commercial video mapping tools. Therefore, in this research, we design a projection mapping software to project multiple video resources easily onto several different flat surfaces at a time.

The proposed software was designed based on the object-oriented method wherein a program consists of a set of classes. Figure 1 depicts a class diagram that includes primary classes of the proposed system. One application class has one-to-many associate relationships with each of the VideoFrame, of VideoPlayer, of Texture, and Animation classes. VideoFrame class is matched with a flat surface including one or less of VideoPlayer and of Texture class members that represent information for one video clip and one still image, respectively. Moreover, VideoFrame has one-to-one associate relationships with of Mesh and of VboMesh classes, which include the initial and current deformed 3D information of a flat surface, respectively. The Animation class defines multiple VideoFrame members displayed at the same time. The application class can have one or more Animation members. Therefore, we can control dynamic display on several flat surfaces in a time sequence.

In the initialization routine, the initial values for members of primary classes are read from the configuration file. The input parameters read from the configuration file are explained in Table 1. Table 2 shows an example of the configuration file to be read in the initialization step. As shown in Table 2, each line has a pair of field name and its value.

After the initialization step is executed, a quadrilateral polygon for each videoFrame is modeled and textured by matching video frame image and texture image. A user can then manually move four polygon vertices, v_0 to v_3, so that the polygon can be deformed with best fit to the rectangular object of the real environment.

2.2 Symmetrical Mesh Refinement for Removing Texture Distortion

Projection mapping is similar to a keystone process that makes a projected quadrangle into a rectangle as shown in Fig. 2(a) [4]. On the other hand, using projection mapping techniques, we project the virtual image onto the object through best fit to its shape as in Fig. 2 (b). In computer graphics, a quadrilateral mesh or a triangular mesh is generally used to represent a 3D mesh model. In other words, a quadrangle and a triangle

Table 1. Input parameters read from the configuration file.

Input Parameters	Description
numVideos	Number of video files to be used
Video#	Filename of the #the video, where # = 1.numVideos
numImages	Number of image files to be used
Image#	Filename of the #the image, where # = 1. numImages
numFrames	Number of VideoFrame instances to be displayed
Frame#_v%_0 Frame#_v%_1 Frame#_v%_2	Initial 3D coordinates of vertex v% of the #the VideoFrame, where # = 1.numFrames and % = 0..3, which means four vertices of a quadrangle
displaySec	Duration of displaying each VideoFrame
numAnimations	Number of Animation instances
numConcurDisplay	Number of VideoFrame instances to be displayed simultaneously
Animation#_%	Index of the %the VideoFrame instance for #th animation

Table 2. Example of configuration file

Page 1	Page 2
%YAML:1.0	Frame2_v0_0: 946
numVideos: 2	Frame2_v0_1: 693
Video1: “CB.avi”	Frame2_v0_2: 0
Video2: “9_video.mp4”	Frame2_v1_0: 612
numImages: 2	Frame2_v1_1: 1485
Image1: “WoodFrame1.jpg”	Frame2_v1_2: 0
Image2: “WoodFframe2.jpg”	Frame2_v2_0: 2087
fadeInSec: 3	Frame2_v2_1: 1795
fadeOutSec: 2	Frame2_v2_2: 0
numFrames: 2	Frame2_v3_0: 2276
Frame1_v0_0: 3050	Frame2_v3_1: 591
Frame1_v0_1: 1342	Frame2_v3_2: 0
Frame1_v0_2: 0	Frame2_ivideo: 2
Frame1_v1_0: 2852	Frame2_iimage: 2
Frame1_v1_1: 2034	displaySec: 800
Frame1_v1_2: 0	numAnimations: 1
Frame1_v2_0: 3609	numConCurDisplay: 2
Frame1_v2_1: 1821	Animation1_1: 1
Frame1_v2_2: 0	Animation1_2: 2
Frame1_v3_0: 3400	
Frame1_v3_1: 1507	
Frame1_v3_2: 0	
Frame1_ivideo: 1	
Frame1_iimage: 1	

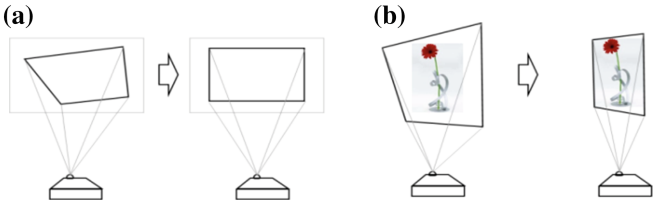


Fig. 2. Comparison between keystone process (a) and projection mapping (b).

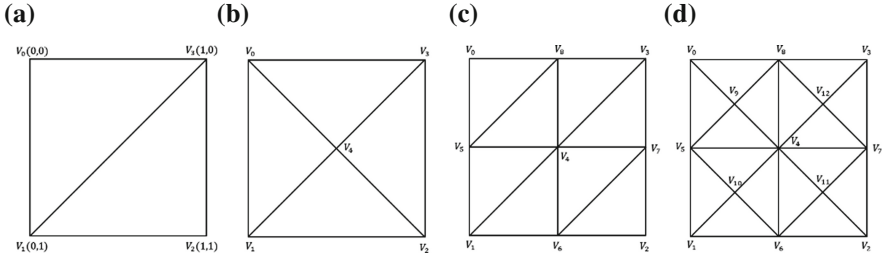


Fig. 3. Different mesh refinement approaches. Mesh refinement (a) using two triangles (b) using four triangles (c) using eight triangles as the most general refinement. (d) Proposed symmetrical mesh refinement.

are primary geometry units to represent a virtual model in computer graphics. In order to display a video clip onto the flat surface of the object, one quadrilateral polygon is first modeled and textured by each frame image of a video clip in sequence. A video frame image mapped on a quadrangle can be easily distorted according to the projection direction since a quadrangle is rendered by two triangles in GPU (Graphics Processing Unit) as in Fig. 3(a).

In this paper, we propose a symmetrical mesh refinement method to reduce video image distortion on a rectangular surface. The mesh refinement types of Fig. 3(b) and (d) are symmetrical up-down-left-right. In 3D modeling, most of the previous refinement algorithms applied the method shown in Fig. 3(c) from the initial mesh of Fig. 3 (a). In the cases of Fig. 3(a) and (c), however, texture image can be severely distorted in the V_0 to V_2 diagonal direction – that is, left-top to right-bottom direction – because changes of vertices on its direction cannot be reflected on changes of vertices on the opposite diagonal direction, that is, left-bottom to right-top direction. In contrast, in the cases of Fig. 3(b) and (d), which belong to the symmetrical refinement type, changes of vertices on a diagonal direction affect the vertices on the opposite diagonal direction, causing all component triangles to be re-textured according to the change of a vertex. Through experiments, we measured the distortion errors of the image projected on a rectangular object after each refinement type of Fig. 3 was applied.

3 Experimental Results

In this work, we implemented the proposed video mapping software using openFrameworks [5] in a Windows 10 environment on the system with Intel® Core i3-5505u@ 2 GHx(4CPU), 8 GB RAM, and Intel HD Graphics 5500 GPU.

In order to measure image distortion errors after the image is projected on the rectangular surface of the real environment using the implemented system, we installed a canvas measuring 130.3 × 97 cm^2 and projected a grid pattern image with 13 vertical lines and 9 horizontal lines. The canvas was installed slantingly with respect to the projection direction to maximize the probability of image distortion. We also used the media artwork image including a large wooden window frame to visualize the effect of image distortion efficiently. Figure 4 shows the canvas installation scene for the experiments.

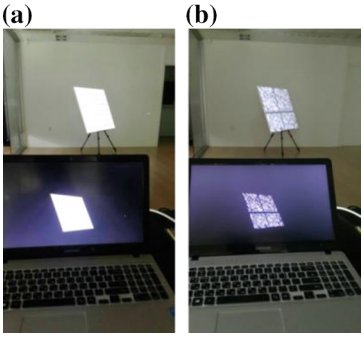


Fig. 4. Distortion experiments using a grid pattern image (a) and a media artwork image (b).

We first captured a screen image including the polygon model deformed with best fit to the canvas shape and textured by the grid pattern image. We recorded the image coordinates of grid points of the 11 × 7 internal grid pattern in the captured screen image. We also computed the ground-truth image coordinates (x_{ij}, y_{ij}) of any internal grid point $[i, j]$ shown in Fig. 5 by linear interpolation among image coordinates of four polygon vertices, v_0 to v_3. Linear interpolation to the horizontal direction was first executed by Eq. 1 and Eq. 2 so that the coordinates of grid points g_{i0} and g_{i8} in Fig. 5 were computed. Linear interpolation between g_{i0} and g_{i8} with respect to index j was then carried out as Eq. 3.

The distortion error was computed by Eq. 4, where (*Xi j, Yi j*) means the observed image coordinates of any internal grid point [i, j] in the captured image. The distortion images after each mesh refinement approach of Fig. 3 are applied are shown in Fig. 6, and the average distortion errors of each refinement approach are presented in Table 3. The proposed software system was used for the projection mapping of a media artwork in the piano concert fused with media art performance. Figure 7 shows the piano concert scene of applying the proposed mapping software.

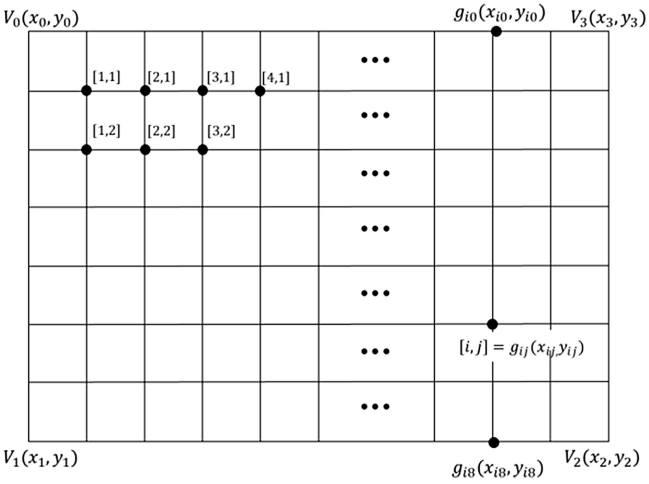


Fig. 5. Grid index $[i, j]$ and image coordinates (x_{ij}, y_{ij}) in the grid pattern image.

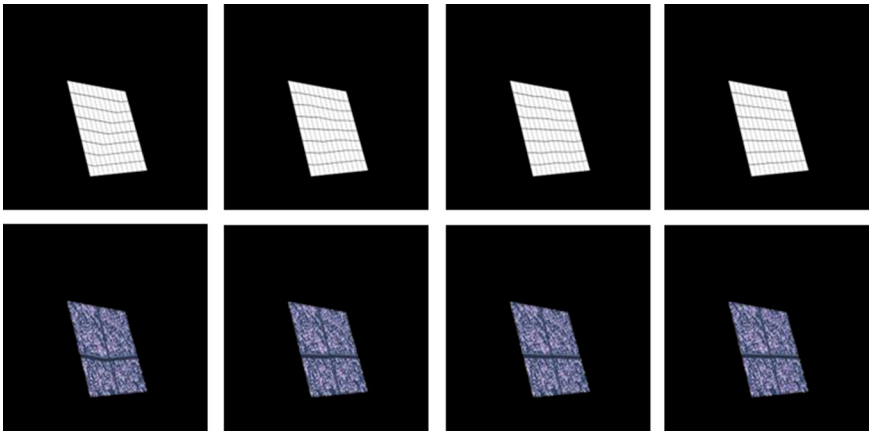


Fig. 6. Image distortion results after each mesh refinement approach of Fig. 3 is applied. (a) tri-2 (b) tri-4 (c)tri-8 (d) tri-16.

Table 3. Average image distortion error (unit: pixel)

Mesh refinement type	tri-2	tri-4	tri-8	tri-16
Average image distortion error	7.673	2.470	1.938	1.010

Fig. 7. Application example. (left) canvas before project mapping, (right) multiple projection mapping scene in the piano concert in collaboration with media artwork

$$x_{i\emptyset} = x_0 \times \left(\frac{12-i}{12}\right) + x_3 \times \left(\frac{i}{12}\right), \quad y_{i\emptyset} = y_0 \times \left(\frac{12-i}{12}\right) + y_3 \times \left(\frac{i}{12}\right), \tag{1}$$

$$x_{i8} = x_1 \times \left(\frac{12-i}{12}\right) + x_2 \times \left(\frac{i}{12}\right), \quad y_{i8} = y_1 \times \left(\frac{12-i}{12}\right) + y_2 \times \left(\frac{i}{12}\right), \tag{2}$$

$$x_{ij} = x_{i\emptyset} \times \left(\frac{7-j}{7}\right) + x_{i8} \times \left(\frac{j}{7}\right), \quad y_{ij} = y_{i\emptyset} \times \left(\frac{7-j}{7}\right) + y_{i8} \times \left(\frac{j}{7}\right). \tag{3}$$

$$E = \sum_{j=1..7} \sum_{i=1..11} \sqrt{(x_{ij} - X_{ij})^2 + (y_{ij} - Y_{ij})^2} \tag{4}$$

4 Conclusion

We presented a projection mapping framework that allows even users with no experience in projection mapping to map multiple video clips onto any flat rectangular surfaces of the real environment easily. Furthermore, through this system, we presented a symmetrical mesh refinement method to reduce image distortion in all projection directions. In computer graphics, a quadrangle and a triangle are primary geometry units to represent a polygonal mesh model that is a display target in projection mapping. Therefore, it is important that images textured on a quadrangle mesh are displayed without severe distortion in order to preserve the visual quality of the projection mapping content. An experiment using a grid pattern image showed that image distortion error was reduced by 95.19 % when a quadrilateral polygon was refined to 16 triangles by the symmetrical mesh refinement. As a future work, we will perform a user test to prove the availability and effectiveness of the proposed system.

References

1. Lee, J.W., Kim Y.J., Kim, D.H.: Real-time projection mapping on flexible dynamic objects. In: Domestic Conference of the HCI Society of Korea, pp. 187–190 (2014)
2. Dumont, T., Ohn, B., Nam, Y.: Projection mapping for flexible surface of object. In: Domestic Conference of the HCI Society of Korea, pp. 155–157 (2014)

3. Bimber, O., Raskar, R.: Spatial Augmented Reality: Merging Real and Virtual Worlds. A.K. Peters, Natick (2005)
4. Song, J.B.: Art and IT – Projection Mapping (2016). http://blog.lgcns.com/1137
5. openFrameworks. http://www.openframeworks.cc

Evolutionary Test Case Generation from UML-Diagram with Concurrency

Seungchan Back[1], Hyorin Choi[1], Jung-Won Lee[2], and Byungjeong Lee[1(✉)]

[1] Department of Computer Science, University of Seoul, Seoul, South Korea
{bsch0lll, yoinoichr2015, bjlee}@uos.ac.kr
[2] Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea
jungwon@ajou.ac.kr

Abstract. To find equality between software products and artifacts, model-based test (MBT) handles specific representations of software requirement. When those conclude concurrency and loop in MBT, it explosively increases a number of paths are applied by existing coverage criteria. Therefore, in this paper, we propose exploration method to avoid path explosion problem, and solution to generate test data automatically using evolutionary algorithm. The result of practical study shows our proposal's efficiency. Testers, who deal with their project through our approach, could find necessary test path. And our approach makes it possible to generate test data according to various test coverage criteria.

Keywords: Test data generation · Model-Based test · Concurrency · UML diagram

1 Introduction

To reduce the cost of software testing, the interest in software testing automation was increased. There are some problems when performing model based testing. One of them is explosively increasing the number of paths that has to be tested when models have concurrency or loop. It is called path explosion. Thus, we propose pair-wise model search for solving path explosion, and test data generation applied evolutionary algorithm from activity diagram. The diagram is one of unified modeling language (UML) models. Proposed model search method severely restricts the number of loop, and explores concurrency path. Also, proposed test data generation uses fitness function including branch distance of B. Korel [1] and approximation level of J. Wegener [2]. Through this paper's proposal, the following is the contributions of our paper.

- Effective test path and test data generation method for testing models concluding loop and concurrency.
- Variety test coverage achievement through just creating target path.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_103

The next chapter introduces the basic approaches and researches related with this paper, and Sect. 3 shows proposed method of searching path and test data generation method. In Sect. 4 proves effectiveness of this paper through experimental result. Lastly, Sect. 5 presents the conclusion with respect to limitations and future works of this paper.

2 Related Work

2.1 Evolutionary Algorithm

The following is the simple conceptual explanation of Evolutionary algorithm. First, generate initial test data set randomly. Next, evaluate suitability of the test data set that are supposed to verify the target part of system. Finally, mutate the set for finding more suitable data with target. P.R. Srivastava [3] proposed an approach that decreases test expenses by generating adaptive test path based on metaheuristic firefly algorithm. Matrix filled by result of calculation of state transition and brightness value and control flow that satisfies the matrix is decided independently then generate test path. This approach satisfies branch coverage with less test cases than conditional algorithm and this would be the contribution of this paper. However, this study does not cover the solution for testing concurrent paths in system under test.

2.2 Path-Wise Testing

Path-wise testing means making and testing test case based on each paths. The paths make it possible to explore a model from start to end. The testing is used mainly in model-based testing and unit testing, and it makes test path satisfying branch coverage as general test goal. Through Multi-Population Genetic Algorithm (MPGA), C. Young [4] proposed path-wise test data generation method. However, MPGA proved effectiveness in comparison with the Single Genetic Algorithm (SGA), it didn't consider loop and concurrency making problem when perform path-wise test data generation. Thus, it is expected to easily occurring path explosion than our approach. N. Esmaeel [5] introduced Path-wise random testing method using adaptive random testing technique. However, it requires another tool to confirm an algebraic representation associated with constraint.

3 Proposed Test Case Generation

Next Sect. 3.1, we propose path search method considering causes like loop and concurrency. And based on approach of Sects. 3.1 and 3.2 shows the test data generation method applying simulated annealing.

3.1 Exploring Method to Generate Target Path

This paper uses Depth First Search (DFS) excepting for certain portions like loop and concurrency, and, in the portion, this paper applied search method of B. SeungChan [6].

For the loop portion, a minimum number of repeat is determined by type of loop, Do-while is 1 and while-do is 2, to assure relation between decision statement and performance statement as it shown in left side of Fig. 1.

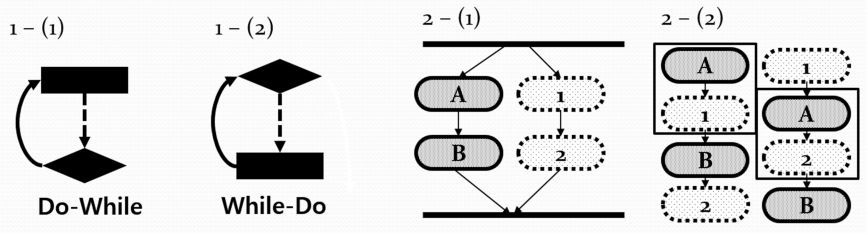


Fig. 1. (Left) Type of Loop and (Right) Example of exploring a concurrency

Concurrency is to operate tasks in threads are processed at the same time. Thus, this portion is divided into previous node and successor node then selected pair of different Threads and DFS is applied. The right of Fig. 1 shows some intermediate path of pair (A, 1) and pair (A, 2).

3.2 Path-Wise Test Data Generation

This chapter proposes test data generation that is finding suitable target path. The proposed method generates test suite randomly, and a test case is evolved by adapting simulated annealing (SA) algorithm. The Table 1 shows the overall algorithm about test suite generation.

TDS is randomly set and it is used as input value to perform test, and F is used to indicate how far target path TP and executed path which is performed by current test data, and F is defined as (1).

$$F = \begin{cases} 0, \text{ if the executed path is equal to the target path} \\ (\text{Approximation Level}) + (\text{Branch Distance}) \end{cases} \quad (1)$$

The top of (1) shows value of F is 0 if executed path is equal to TP. In this case, TP is removed from the set of to be tested TP. In other words, current test data has a capability to test TP. When executed path isn't equal to TP, F of the bottom of (1) uses Approximation Level (AL) and Branch Distance (BD) which are widely used in other works [1–3]. AL indicates nesting level is the number of the decision between executed path and TP, important thing is that gap between ALs is greater than gap between BDs. BD represents distance to change value of decision for matching up with TP. If the value of decision is T, BD represents distance for becoming F.

Table 1. Test suite generation algorithm

Input Target Path *TP = { TP1, TP2, … , TPn-1, TPn }*

```
1    Set an initial test data set TDS = { TDS1, TDS2, … , TDSn-1. TDSn }
2    Set fitness function f
     Evaluate TDS by adapting f
3   for i ← 1 until n do
4     for j ← 1 until n do
5       fv = f(TPi, TDSj)
6       min(TPi) ← Find minimum value fv
7     End for
8     Θ = min(TPi)
9     mTDS ← TDS related with Θ
10    If( Θ = 0 )
11      Then mTDS is added to TS,
12    Else
13      SA( mTDS , TPi )
14    End if
15    TPi is deleted from TP
16  End for
```

Output Test Suite *TS*

The Loop beginning on line 4 examines whether randomly generated TDS is suitable to TP using F. And in line 14, if TP does not have any other suitable TDS, the closest executed path is selected and it is advanced through SA. Detailed algorithm of SA is shown in Table 2.

Table 2. Simulated Annealing (SA) algorithm applied our approach

Input SA(*TDS, TP*)

```
1    Set an initial temperature t,
2    Set V={TDS | v1,v2,…,vn},
3    While (AL ≠0)
4      Find effectible value set of V, EV
5      Set for each of EV's Neighborhood range
6      While (BD(TDS, TP) ≠ 0)
7        Select a new solution sTDS ∈ Neighborhood(TDS)
8            θ = BD(sTDS, TP) - BD(TDS, TP)
9        if ( θ ≤0) Then TDS ← sTDS
10       Else x = a uniform random value in range[0..1]
11         if (x < exp(- θ /t)) Then TDS ← sTDS
12         End if
13       End if
14       Cooling temperature t
15     End While
16   End While
```

Output Suited *TDS*

First and second line of Table 2 shows t value (temperature) which is used in SA method then declares a set of variable V held by the TDS received as a parameter value. From line 3 to line 16 represent finding TDS that satisfy TP. If Approximation Level (AL) value is 0 it means TDS has reached its TP. Line 4-5 find a value that changes Branch Distance (BD) value among V set. Neighborhood range is scope of TDS value which is possible to be changed one time. From line 6, there is a while loop that is for change branch distance value in TDS to 0.

New solution TDS that is generated by changing value as Neighborhood is chosen, then compare TDS with BD and if the comparison value θ is negative, substitute TDS with sTDS, if it is positive takes same action as negative value when it is higher than certain probability. Finally, when all AL and BD value are 0 then new TDS value which is possible to find TP is returned.

4 Experimental Result

In this chapter, we show the effectiveness of our approach through experimental results. The tool used of this paper's experiments is Optimization Tool Box which provides a variety of evolutionary algorithms and convenient appliance. The model is '*validDate*' activity diagram shown below Table 3 which is used in typically to verify the performance of the evolutionary algorithms.

Table 3. A Result of the experiment in validDate

Activity diagram	Number of branches	Number of variables	Number of activities	Concurrency threads
Valid date	11	3	10	3
	Covered (%)	The number of test case	Description	
Basis path Coverage	100	12	All transition explores more than once.	
Simple path Coverage	100	360	Basis path, all concurrency path and loop path is performed once.	
Our proposed Coverage	100	16	Concurrency path is searched according to our technique, and loop statement is performed 1–2 times, depending on the type.	

5 Conclusion

This paper presents an approach for contributing automated software testing that considers concurrency. We improved other existing approaches by searching feasible paths of models and generating test cases which are generated for each path. This approach decreases effort for manual testing by finding suitable test data based on evolution algorithm. It is possible to generate sets of test data meet full coverage when

path is generated. However, we has only performed only in the activity diagram, thus, the implementation of a different behavior model is necessitated such as state chart diagram and sequence diagram. In the future, we are planning to solve problems that mentioned before and develop a system that supports these approaches.

Acknowledgments. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030504), and by Seoul Creative Human Development Program funded by Seoul Metropolitan Government (No.CAC1510).

References

1. Korel, B.: Automated software test data generation. IEEE Trans. Softw. Eng. **16**(8), 870–879 (1990)
2. Wegener, J., André, B., Harmen, S.: Evolutionary test environment for automatic structural testing. Inf. Softw. Technol. **43**(14), 841–854 (2001)
3. Srivasstava, P.R., Mallikarjun, B., Yang, X.S.: Optimal test sequence generation using firefly algorithm. Swarm Evol. Comput. **8**, 44–53 (2013)
4. Chen, Y., Yong, Z.: Automatic path-oriented test data generation using a multi-population genetic algorithm. In: Proceedings of Fourth International Conference on Natural Computation, vol. 1, pp. 566–570 (2008)
5. Nikravan, E., Farid, F., Saeed, P.: Enhancing path-oriented test data generation using adaptive random testing techniques. In: Proceedings of Second International Conference on Knowledge-Based Engineering and Innovation, pp. 510–513 (2015)
6. Seungchan, B., Hyorin, C., Byungjeoung, L., Jung-Won, L.: A test scenario generation method from activity diagram with concurrency. In: Proceedings of Korea Conference on Software Engineering, pp. 44–45 (2016)

Evaluating the Effectiveness of the Vector Space Retrieval Model Indexing

Jung-Hoon Shin[1], Mesfin Abebe[2], Cheol Jung Yoo[1], Suntae Kim[1(✉)], Jeong Hyu Lee[1], and Hee-Kyung Yoo[3]

[1] Department of Software Engineering, CAIIT, Chonbuk National University, 567 Baekje-daero, Deokjin-gu, Jeonju-si, Jeollabuk-do, Republic of Korea
{shinjh, cjyoo, stkim, jhlee25}@jbnu.ac.kr
[2] Department of Computing, Adama Sience and Technology University, Adama, Ethiopia
mesfinabha@gmail.com
[3] Department of Computer Science and Engineering, Kangwon National University, JoongAng Ro 346, Samcheck-Si, Kangwon Do, Republic of Korea
hkyoo@kangwon.ac.kr

Abstract. Modern information retrieval activities are supported with software systems that facilitate the users' information searching. Information retrieval systems are significantly improved in the past few decades. Now days, there are three types of retrieval models: *Boolean*, *Vector Space* and *Probabilistic*. In this study, we examined the vector space model where documents and queries are represented as vectors. We conducted a number of experiments on the i*ndexing* technique of the vector space model to quantitatively describe the effectiveness of the techniques using *Lemur* Toolkit. The result indicates that stop word removal and steaming techniques improve the quality of the index terms.

Keywords: Information retrieval · Vector space model · Indexing · Similarity function

1 Introduction

Information Retrieval (IR) is the activity of looking for relevant information from a large collection of information bearing items such as documents, images, music, videos and others [1]. Modern information retrieval activities are automated with the help of software systems and they are used with digital libraries, recommender systems, and search engines etc. *Pulling* and *Pushing* modes are the two ways in which information retrieval systems (*IRSs*) retrieve relevant documents [2]. In the *pulling* mode, the users are responsible to initiate the information retrieval process; nevertheless, the system starts the retrieval process in the *pushing* mode. This study discussed information retrieval from the perspective of the pulling mode.

A typical information retrieval system has a number of modules; however, the following three are the common components of an IR system [3]:

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_104

- ***Document database:*** characterizes the document collection in the form of index terms that simplify the information retrieval activity.
- ***Information needs (Queries):*** represent the user information need as a query to retrieve the relevant documents or information.
- ***Matching and Ranking mechanism:*** evaluates the similarity between the documents and the user information need to determine the relevant documents.

In this study, we quantitatively evaluated the vector space model indexing technique. The study covers the main activities (pre-processing techniques) of the model during the document indexing. The rest of the study is organized as follows: Sect. 2: gives an overview of the vector space model. Section 3: explains the experimental activity. Section 4: presents and discusses the result of the study. Finally, Sect. 5: conclusions and proposes a future study area.

2 Overview of Vector Space Model Information Retrieval

Information retrieval systems (*IRSs*) are used to retrieve information from unstructured or semi-structured document collection based on the users' information need [4]. The vector space model is an algebraic retrieval model that represents the documents and the queries as vectors [5]. Therefore, both the document collections and information needs (queries) require to be transformed into a bag of words (BoW) or indexes. The vector space model was criticized a lot for its rough approach at the early stage of it inception [6], but currently the model improves in many direction.

The vector space model goes through a lot of changes from its earlier stage to minimize its limitations and increase its retrieval effectiveness. Normally, there are many studies on the indexing, weighting and matching of the model, though most of them do not describe the benefits in quantitative manner [7]. The vector space model has three main elements such as *indexing*, *term weighting* and *matching*; however, in this study we focused on the indexing section of the model.

Indexing is the process of converting the documents into a bag of words to represent the document collection in a data structure that is easy and simple for retrieval [2]. Many of the tokens in the document collection are not significant to describe the content of the documents. The vector space model uses different pre-processing techniques to reduce the non-significant words, so that the size of the index terms can be decreased to a manageable size. Table 1 shows the commonly used pre-processing techniques of a vector space model.

Table 1. Commonly used pre-processing techniques

No.	Pre-processing	Description
1	Tokenization	Chops down the documents into tokens (words) using delimiters like whitespace
2	Stop word removal	Removes commonly appeared stop words such as a, the, it etc.
3	Stemming	Strip off suffixes and prefixes to change the token to root word
4	Indexing	Builds an inverted index data structure to provide fast searching

- ***Tokenization:*** is a strategy to whittle down the document collection into terms (tokens) using delimiter such as whitespace and punctuations [8]. The end products of tokenization are a sequence of characters called *tokens*.
- ***Stop word removal:*** is the removal of non-significant terms from the tokens to reduce the number of index terms [9].
- ***Stemming:*** is the extraction of the suffixes and prefixes to minimize the word variation (*derivational forms*) using the root (*stem*) words [10].
- ***Inverted index:*** is the activity of building a data structure to support fast retrieval [5]. It consists of a collection of posting lists and index terms.

3 Experiment on the Pre-processing Techniques

In this study, we investigated the indexing process of a vector space information retrieval model. The first subsection describes the experiment organization and the second subsection explains the result of the experiment.

3.1 Organizing the Experiments

The experiment includes the *stop word removal* and *stemming* (*Krovetz* and *Porter*) techniques of the vector space model. We used the Lemur information retrieval toolkit and three groups of document collections for the experiment. We investigated the model indexing effectiveness without *stop word removal*, without *stemming*, with *stop word removal* and with *stemming*. The study followed the following steps:

- Formatting the documents in *TREC* format for the study (*60 documents*)
- Preparing the stop word list.
- Indexing the sample documents using *Lemur* tool kit by applying different pre-processing techniques.
- Analyzing and discussing the result.

3.2 Pre-processing Techniques (Stop Word Removal and Stemming)

The *Lemur* toolkit creates a bag-of-words to represent the document collection and to store the index terms location of the document collection. The terms location information specifies standard information such as document lengths, document frequency and term collection frequency. The sample documents are extracted and wrapped in TREC-style wrappers using <DOC>, <DOCNO> and <TEXT> tags. In addition, we used a list of English language stop words *(301 words).* The stop words list is defined within a <stopper> tag and each stop word enclosed within a <word> and </word> tags. The document collection consists of a total of *60* documents that have different content and size. The documents are splitted into three groups to conduct each indexing experiment three times for validation purpose.

Table 2. Pre-processing techniques used in the experiment

File	Applied techniques	Description
I1	Without stop words list	The documents are indexed without using any stop word list and stemming technique
I2	With stop word list	The documents are indexed using a stop word list, but without stemming
I3	Krovetz stemming without stop words	The documents are indexed by applying Krovetz stemming without stop word removal
I4	Portar stemming without stop words	The documents are indexed by applying Portar stemming without stop word removal
I5	Krovetz stemming	The documents are indexed using stop word and Krovetz stemming
I6	Portar stemming	The documents are indexed using stop word and Portar stemming

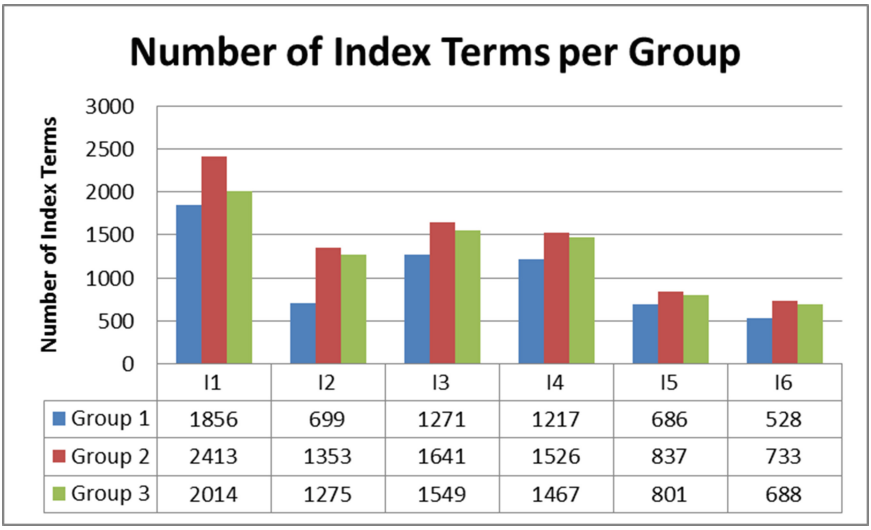
Each group documents are indexed into six different index files by applying various preprocessor techniques as shown in Table 2. The index files are constructed by running the *Indr Indexer* over the sample documents.

The *index files* are different in size and word variation due to the techniques we applied to create the six index files. The index terms are decreased significantly in size and type from *index file one* to *index file six* (*I1 to I6*). The following bar graph illustrates the index size of the six files using the three groups of documents.

4 Result and Discussion

The pre-processing techniques; stop word removal and stemming are applied on all the eighteen (18) index files. The sixty (60) documents are grouped into three clusters as shown in Fig. 1. Each group consists of twenty documents that are used to establish the six index files (I1, I2, I3, I4, I5 and I6). The size of the index files shrunk in size and type of index words as we applied more and more pre-processor techniques.

Applying the stop word removal and stemming techniques reduced the size and the quality of the index terms. However, vector space model problems such as missing semantic and syntactic information are not improved with these techniques. Representing the document collection with few but unique terms can improve the effectiveness of the retrieval process. As shown in Table 3, using the stop word removal technique reduced the total number of the index term by 47 %. The Krovetz and Portar stemming techniques are condensed the index file size by 29 % and 33 % respectively. Similarly, by combining the stop word removal and stemming techniques the size of the index words reduced by 63 % in the case of Krovetz stemming and by 69 % in the case of Portar stemming method.



	I1	I2	I3	I4	I5	I6
Group 1	1856	699	1271	1217	686	528
Group 2	2413	1353	1641	1526	837	733
Group 3	2014	1275	1549	1467	801	688

Fig. 1. Numbers of index terms using the three groups of documents

Table 3. Reduction in the index files due to the stop word removal and stemming

File	Group 1	Group 2	Group 3	Average	Reduction in %
I1	1856	2413	2014	2094	00.00 %
I2	699	1353	1275	1109	47.00 %
I3	1271	1641	1549	1487	29.00 %
I4	1217	1526	1467	1403	33.00 %
I5	686	837	801	775	63.00 %
I6	528	733	688	650	69.00 %

5 Conclusion

In this paper, we study the indexing techniques of the vector space information retrieval model such as stop word removal, and stemming. The examination was conducted using the Lumer toolkit and a collection of documents. The indexing experiment is done on three groups of documents. We created six different index files with each group documents using various preprocessors techniques. The experiment revealed the significant impact of each technique in a quantitative way. The evidence can encourage researches to work more on the area to improve the overall quality of the retrieval model. In the future, we would like to extend the study by examining the index files with different ranking and matching techniques of the Lumer Toolkit and investigate the effect on the precision and recall of the information retrieval.

Acknowledgement. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030503).

References

1. Singh, J.N., Dwivedi, S.K.: A comparative study on approaches of vector space model in information retrieval. In: International Conference of Reliability, Infocom Technologies and Optimization (2013)
2. Khan, J.A.: Comparative study of information retrieval models used in search engine. In: IEEE International Conference on Advances in Engineering and Technology Research (2014)
3. Ahmed, F., Nurnberger, A.: Literature review of interactive cross language information retrieval tools. Int. Arab J. Inf. Technol. **9**(5), 479–486 (2012)
4. Zuo, J., Wang, M., Wano, J., Luo, W.: Information retrieval model combining sentence level retrieval. In: International Conference on Asian Language Processing (2013)
5. Fang, H., Tao, T., Zhai, C.: Diagnostic evaluation of information retrieval models. ACM Trans. Inf. Syst. **29**(2) (2011)
6. Raghavan, V.V., Wong, S.K.M.: A critical analysis of vector space model for information retrieval. J. Am. Soc. Inf. Sci. **37**, 279–287 (1986)
7. Kolda, T.G.: Limited-memory matrix methods with applications, Applied Mathematics Program. University of Maryland at College Park, pp. 59–68 (1997)
8. Dong, H., Hussain, F.K., Chang, E.: A survey in traditional information retrieval models. In: IEEE International Conference on Digital Ecosystems and Technologies, pp. 397–402 (2008)
9. Al-Dubaee, S.: New information retrieval model. In: Science and Information Conference, pp. 819–826 (2014)
10. Jivani, A.G., et al.: A comparative study of stemming algorithms. Int. J. Comput. Appl. **2**(6), 1930–1938

Active Tracking Strategy with Multiple Cameras in Large Areas

Sangjin Hong[1] and Nammee Moon[2(✉)]

[1] Stony Brook University, New York, USA
[2] Hoseo University, Asan, South Korea
mnm@hoseo.edu

Abstract. This paper presents a distributed multiple camera collaboration strategy for active tracking in large areas. Their collaboration utilizes a sector-based mechanism in which the visual correspondences among different cameras are determined. In order to emulate the master-slave operations, the proposed method combines the local data to construct the global information. Based on the global information, the loads are evenly distributed to avoid tracking misses due to the dynamic behaviors of the objects. The trajectories of all objects visible in the local cameras are estimated for actual tracking and the estimated dynamics are used for scheduling of the cameras. The active tracking triggering timing is carefully chosen to maximize the overall monitoring time for general surveillance operations.

Keywords: Active tracking · Surveillance systems · Camera collaboration

1 Introduction

Multiple cameras are incorporated into the sensor networks to visually monitor the objects in many areas such as airports and industrial complexes for public safety purposes. In order to minimize the issue of limited resolution, many interesting works on tracking with a single active camera have been proposed [1–3]. Most works on active tracking utilizing single camera were focused on coping with the slow translation speed of the cameras. In order to cover large areas as well as densely populated objects, many of these single camera active tracking systems have been extended to networked multiple camera active tracking schemes [4–8].

In many multiple camera active tracking systems, the master-slave approaches were often used where the master covers the entire area and the slave cameras perform active tracking of a specific object based on the priority-based or the round-robin based scheduling [9–11]. With the master-slave approach, the process of the camera scheduling and their coordination is straightforward but resource over-utilization can happen. Moreover, scalability of the system is not easily achieved without having multiple master cameras.

Hence, distributed approach for multiple camera active tracking may be preferable to maximize the camera utilization and achieve scalability of the system. However, distributed approach has loose coordination among the cameras and the system may not support dynamically varying object distribution densities. Thus, in covering a large

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_105

common region with distributed multiple cameras, the overall correspondence information of objects among cameras is necessary in order to minimize the redundant active tracking.

In this paper, a distributed multiple camera collaboration strategy for multiple camera active tracking in a large area is proposed. The proposed mechanism utilizes the sector-based representation between the cameras so that the correspondence among different cameras is determined. In order to cover a large area, the master-slave operations are emulated by combining local object profiles to construct the overall global coverage information. The global coverage information utilizes the estimated trajectories of all objects visible by the local cameras where the information is used to select an object for the active tracking. The process of load balancing among cameras to evenly distribute the objects is incorporated into the mechanism to avoid tracking misses due to the dynamic behaviors of the objects. The active tracking triggering timing is carefully chosen to synchronize the cameras and maximize the overall monitoring coverage for general surveillance operations.

2 Basic Sector Distributed and Collaboration Strategy

2.1 Sector Distribution for Multiple Cameras

Figure 1 illustrates a large area covered with a set of multiple PTZ cameras where the coverage of each camera depends on the PTZ parameters. By varying the PTZ values, various view angles are possible. The cameras are placed in so that they could cover the entire global view with some possible coverage overlaps. The global area is divided into a set of sectors and each sector is uniquely indexed. The coverage of each camera is represented by a set of the global sectors that the camera monitors. The sector based representation is used to establish the correspondences between multiple cameras.

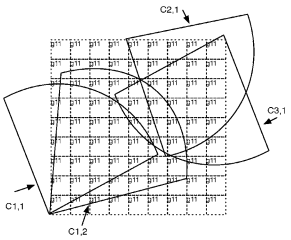
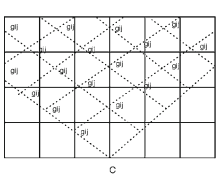
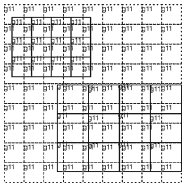


Fig. 1. Illustration of the cameras covering the different global area depending on the PTZ parameters. Each camera has multiple view positions.

Figure 2(a) illustrates the correspondence between the global view and local camera view. The view of the camera is also represented as a set of sectors. Initial objective of using the sectored based approach is to obtain a mapping coverage solution to maximize the global view coverage (i.e., maximizing coverage of objects). Because the size of global sectors projected on the local sector may not correspond to the sizes



(a) Local View Perspective



(b) Global View Perspective

Fig. 2. (a) Illustration of that the projection of global sectors to camera local view may have a distortion (the local view of camera is shown). (b) Illustration of the sector correspondence between the camera sectors and the global sectors.

of the local sectors, multiple global sectors may be mapped to multiple local sectors as illustrated in Fig. 2(b).

Two main data structures are maintained for efficient mapping process. Figure 3(a) shows the data structure for the global sector information. For each global sector, the corresponding local sectors for each camera position are maintained. Multiple local camera sectors may be mapped to a single global sector (or vice versa). In addition to this global sector information, a data structure is maintained for each camera as shown in Fig. 3(b). For each camera and its position, the list of corresponding global grid list is provided for the sector. The correspondence between the camera local sectors to the global sectors is established through the mapping data structure maintained for each camera. With the mapping table for each camera, the global to local view relationship is clearly specified. Also, by checking which local sectors of different cameras contain the same global sector indices, which means that they cover the same global sectors, the correspondence between different camera views could be easily achieved.

Grid Index	Camera i			...	Camera i		
	Position i	...	Position k	...	Position i	...	Position k
g11	s11, s23	...	s11, s23	...	s11, s23	...	s11, s23
g12	g12	...	g12	...	g12	...	g12
g12	g12	...	g12	...	g12	...	g12
...	...	...	...	...	...	...	...
gkm	gkm	...	gkm	...	gkm	...	gkm

(a) Global View Based

Sector Index	Position i Grid List	...	Position k Grid List
s11	g11, g23	...	g11, g23
s12	...	...	...
s12	...	...	...
...	...	...	...
skm	...	...	...

(b) Local View Based

Fig. 3. (a) Data structure maintaining the global sectors to the local sectors for the individual placement position of all cameras. (b) Data structure maintaining the local camera sectors to the global sectors for individual placement position.

2.2 Solution for Load Balancing

As discussed in the previous section, the mapping method generates many possible solutions. Moreover, the mapping method discussed previously does not consider the object distribution and their dynamic. When selecting a mapping solution, the load balancing must be considered at the same time. The main objective of the load

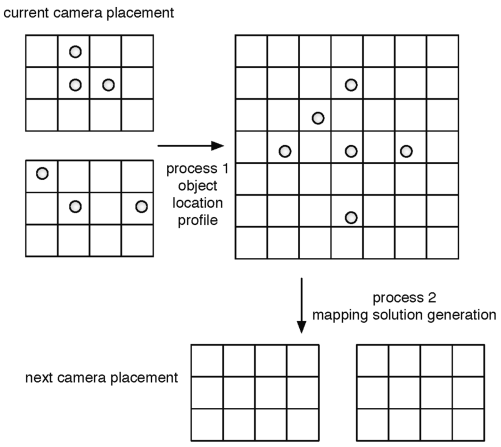


Fig. 4. Estimating the object distribution on global view using projection of the locally detected objects. Form the global estimation, the mapping solution as well as load balancing is achieved.

balancing is to distribute the object evenly among the cameras so that the total time it takes to cover the objects is evenly distributed.

Figure 4 illustrates the strategy for estimating the objects distribution on the global view by projecting the locally detected objects. (For simplicity, the discussion considers the case with perfect projections, that is, the local views have no distortion in terms of global views. This will not affect our algorithms and the results.) The objects detected locally on each camera are projected using the pre-computed relationship between the global area and local view. This method is necessary in order to determine the object distribution, hence proper solution selection for load balancing. Since this method projects all objects that are locally detected, the method cannot estimate the area, which is not being covered by the camera. In the process, depending on the relative size of the grid and sectors as well as degree of overlapping, the object distribution estimation may not be perfect. Moreover, the object distribution is obtained only with the viewable ranges of the cameras. The objects that are not visually covered by the camera are not considered in the estimation process. Hence, normal coverage of the entire global view is important for obtaining the accurate estimation of the object distribution.

After load balancing, the system will then perform sector scanning and object selection for active tracking. Each camera scans through their local sectors sequentially. Whenever there is a detected object in a certain sector, the camera selects the object and initiates the zooming process. Since object tracking is not possible during the zooming process, whenever there is a sector with multiple detected objects, the scanning process randomly selects one of the objects and initiates the zooming process. The cameras also keep counters for each sector indicating the number of the objects within the sectors. Each time the camera scans and tracks the sectors, the counter for that sector will decrease by one. Since the two objects still have the same probability to be selected during the next scanning iteration, possible missed and redundant detections are likely to occur.

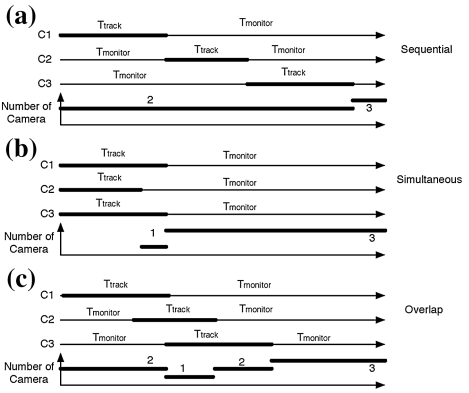


Fig. 5. Illustration of cameras collaborations for three different start time of the active tracking process. (a) Sequential, (b) Simultaneous, (c) Partial overlap.

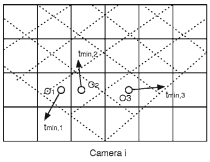
The discussion of the single camera scanning and selection strategy can be extended to the case for the multiple cameras from the timing characteristic and the overall coverage perspectives. Figure 5 illustrates the active tracking collaboration strategies for different start time of the active tracking process. The parameter $T_{monitor}$ represents the time duration the camera is monitoring the area and the parameter T_{track} represents that the camera is performing the active tracking for all objects assigned to the cameras. The figure assumes that the cameras C_1, C_2, and C_3 cover five objects, four objects, and four objects, respectively. The first strategy shows that the camera is sequentially performing the active tracking. During the duration of active tracking operation, the number of cameras that are participating in overall monitoring is limited to two. The second strategy is suitable when all of the cameras are performing active tracking simultaneously. Initially, the number of cameras covering the area is zero but the number increases to three when the active tracking is completed. The last strategy overlaps the active tracking activities and the coverage is also illustrated.

3 Dynamic Sector Distribution with Load Balancing

3.1 Consideration of Object Dynamics

The dynamics of the objects on the local camera view have great influence on the scanning and selection of the system. Consider three objects illustrated in Fig. 6(a). While all of these objects may move with the same speed on the global view, the projected speed of these objects are different depending on the view angle of the camera. In this particular case, the object O_3 may leave the view before the object O_1.

Without proper handling of the object dynamic, the system may miss many objects for the active tracking operation. In the proposed approach, a set of parameters is characterized. The parameter $t_{min,i}$, for each object i, specifies the minimum time required in order for the object to leave the coverage view of the camera. The values are



(a) Object Trajectories on Local View

Object ID	Global Position	Camera 1 tmin	Camera 2 tmin	...	Camera k tmin
O1	g1,1	0.1	-	...	3
O2	g1,1	0.7	-	...	2
O3	g1,1	60	0.3	...	-
...	...	...	...	...	...
ON	g1,1	20	-	...	-

(b) Data Structure for $t_{min,\,i}$ for Object i

Fig. 6. (a) Illustration of the t_{min} depending on the objects location on the local view. The time varies significantly depending on the direction of movement. (b) The table storing the estimated t_{min} for each object for all cameras.

estimated by the local cameras for all objects, and these values are tabulated by the server. The data structure maintained by the server is illustrated in Fig. 6(b).

3.2 Synchronization of Active Tracking with Constraint

With all the priority information mentioned above, cameras still have to exchange information with the server in order to collaboratively increase the total coverage and to be effectively scheduled. The local camera sends the information or possible ordering to the server for deciding the active tracking instances. The reason for this is to avoid multiple cameras processing the active tracking at the same time reducing the overall coverage.

Figure 7 illustrates the ordering of objects received by the server. Each object i has its $t_{min,i}$ and as long as the object is selected to be tracked before the given $t_{min,i}$, the object is safely tracked. Beyond the $t_{min,i}$, the object will not be covered by the active tracking. Hence, the cameras constantly provide the estimated time of exits of the remaining objects. Based on these, the load balancing as well as active tracking

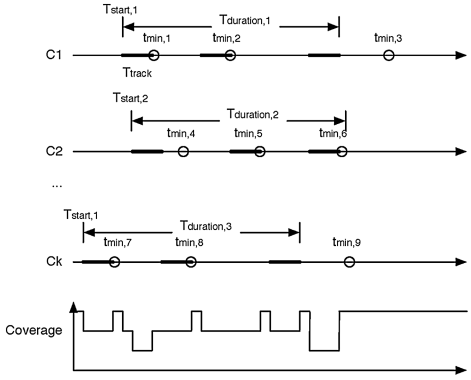


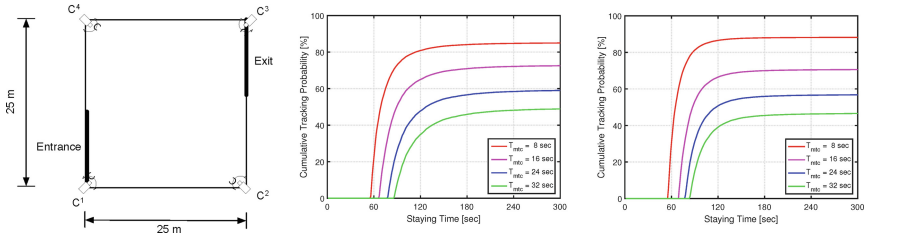
Fig. 7. The ordering information received by the server. The camera sends the T_{exit} of all remaining objects to the server. The overall coverage is influenced by the active tracking synchronization.

synchronization is performed. The server individually interfaces with the cameras one at a time with $T_{start,i}$ signal for the camera i with the duration of active tracking $T_{duration,i}$ for the camera i. This will initiate the receiving camera to perform active tracking for the specified duration. Within the duration, multiple objects may be tracked. The reason for this individual synchronization is to maintain the monitoring coverage by the cameras. Hence, the active tracking synchronization is performed with the number of cameras participating in the monitoring as a constraint.

One issue may arise during the active tracking, many objects may move and the object may not be able to cover within the given $t_{min,i}$, which means that not enough to handle the current load of objects. When the timing is violated, some of the objects will not be tracked.

4 Evaluations

In this section, the effectiveness of the proposed method is evaluated with an environment as shown in Fig. 8(a). The entrance and exit are indicated in the setup. Multiple cameras with a finite set of PTZ parameters and the finite active tracking time $T_{track,min}$ are used and the number of cameras is varied in the evaluation.



(a) Setup (b) Translation Time = 0.5 *sec*, C^2, C^4 (c) Translation Time = 0.25 *sec*, C^2, C

Fig. 8. (a) The simulation setup using four cameras in a large area where each camera may have different views. The miss active tracking performance as a function of T_{mtc}. Two different sets of cameras are compared. The value of the camera translation time is set to (a) 0.5 s (b) 0.25 s. Camera Positions 2 and 4.

Figure 8(b) illustrates the percentage of the objects with missed active tracking. The percentage depends on the amount of time the objects stayed within the area as well as the camera positions. As shown in the figure, the smaller the value of T_{mtc} is, the lower the miss tracking. Since the objects were diffused from one location, the performance depends on the camera position. Figure 8(c) illustrates the same evaluation with the faster translation time, 0.25 s. It is clear that when the cameras are limited, the performance with the faster translation camera is highly desirable. In both figures, 100 % cumulative tracking probability is not possible since some of the object which entered with very short staying time is permanently missed by the system.

5 Conclusions

In this paper, a distributed multiple camera collaboration strategy for multiple camera active tracking in a large area is proposed. The proposed method utilizes a sector-based mechanism in which the correspondence among different cameras is determined. By emulating the master-slave operations through the combination of the local information to construct the global information, the loads are evenly distributed to avoid tracking misses due to the dynamic behaviors of the objects. The performance of the proposed mechanism is evaluated with a densely populated situation. When the resources are limited, the proposed scheduling mechanism tries to maximize the active tracking coverage as well as the normal coverage. It is also demonstrated that faster translation time of the cameras generates better performances.

References

1. Cho, S.H., Nam, Y., Hong, S., Cho, W.: Sector based scanning and adaptive active tracking of multiple objects. TIIS **5**(6), 1166–1191 (2011)
2. Dinh, T.B., Vo, N., Medioni, G.: High resolution face sequences from a PTZ network camera. In: IEEE International Conference on Automatic Face Gesture Recognition and Workshops (FG 2011), pp. 531–538 (2011). doi:10.1109/FG.2011.5771454
3. Varcheie, P.D.Z., Bilodeau, G.A.: People tracking using a network- based ptz camera. Mach. Vis. Appl. **22**(4), 671–690 (2011). doi:10.1007/s00138-010-0300-1
4. Zhao, L., Liu, L.F., Wang, Q.Y., Li, T.J., Zhou, J.H.: Moving target detection and active tracking with a multicamera network. Discrete Dyn. Nat. Soc. **2014**, 11, Article ID 976574 (2014)
5. Huang, C.-M., Fu, L.-C.: Multitarget visual tracking based effective surveillance with cooperation of multiple active cameras. IEEE Trans. Syst. Man Cybern. Part B Cybern. **41**, 234–247 (2011)
6. Hoffmann, M., Wittke, M., Bernard, Y., Soleymani, R., Hahner, J.: DMCTRAC: distributed multi camera tracking. In: ACM/IEEE International Conference on Distributed Smart Cameras (2008)
7. Krahnstoever, N., Yu, T., Lim, S.-N., Patwardhan, K., Tu, P.: Collaborative real-time control of active cameras in large scale surveillance systems. In: Workshop on Multi-camera and Multi-modal Sensor Fusion (M2SFA2) (2008)
8. Ward, C., Naish, M.: Scheduling active camera resources for multiple moving targets. In: Canadian Conference on Electrical and Computer Engineering (CCECE 2009), pp. 528–532 (2009). doi:10.1109/CCECE.2009.5090187
9. Costello, C.J., Diehl, C.P., Banerjee, A., Fisher, H.: Scheduling an active camera to observe people. In: Proceedings of the ACM 2nd International Workshop on Video Surveillance & Sensor Networks (VSSN 2004), pp. 39–45. doi:10.1145/1026799.1026808
10. Qureshi, F., Terzopoulos, D.: Surveillance camera scheduling: a virtual vision approach. Multimedia Syst. **12**(3), 269–283 (2006)
11. Krahnstoever, N., Yu, T., Lim, S.N., Patwardhan, K., Tu, P.: Collaborative real-time control of active cameras in large scale surveillance systems. In: Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, Marseille, France (2008). http://hal.inria.fr/inria-00326743

A Survey and Design of a Scalable Mobile Edge Cloud Platform for the Smart IoT Devices and It's Applications

Yeongpil Cho[1], Yunheung Paek[1], Ejaz Ahmed[2], and Kwangman Ko[3(✉)]

[1] Electrical and Computer Engineering, Seoul National University, Seoul, Korea
[2] Department of Computer System and Technology, Malaya University, Kuala Lumpur, Malaysia
[3] School of Computer and Information Engineering, SangJi University, Wonju, Korea
kkman@sangji.ac.kr

Abstract. Mobile Edge Computing offers real time RAN information (like network load, user's location) to the application developers and content developers. These real time network information are used to provide context aware services to the mobile subscribers, thereby enriching user's satisfaction and improving Quality-of-Experience. Mobile edge computing platform increases the edge responsibility and allows computation and services to be hosted at the edge, which reduces the network latency and bandwidth computation of the subscribers. In this research paper, we designed the scalar mobile edge cloud platform and it's appropriate edge applications, such as 360° panorama image processing, which has a special characteristics and challenges to extend an edge servers on edge cloud by the subscribers demands. This research paper challenges capable to overcome the static al constrains of edge serve capacities and supports flexible computing facilities.

Keywords: Mobile Edge Computing · Communication offloading · Resource scheduling · Energy consumption management

1 Introduction

Internet-of-Things (IoT) paradigm enabled the resource-constrained devices to be interconnected through Internet [1]. But, many of these edge devices are embedded with low processor and storage capacity. To overcome above scenario in Mobile Cloud Computing (same scenario as IoT paradigm), many techniques such as cyber foraging [2, 3] or computational offloading [4, 5] have been proposed where edge device offloads some computation to the remote resourceful cloud, thereby saving processing power and energy. However, offloading computation to the public cloud may involve long latency for data exchange between the public clouds and edge device through the Internet. To overcome above problem, cloudlet based offloading is proposed where mobile devices offload computational to the less resourceful server near the uses proximity accessible using Wi-Fi access point.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_106

In this paper, we proposed and designed a scalable-edge cloud platform which an edge device subscribers could flexible extends cloud edge server to edge cloud and could application services specific to edge devices requirements in current static environments. And then, we verified the overall system completions and performances using the appropriate edge device application development. A scalable edge cloud platform composed of edge cloud controller and edge device controller. For the appropriate experimental edge device application, 360° panorama to plain view translation, moving object recognition, moving object tracking, we'll develop the 360° panorama image processing application that offloaded to rich resource computing server. Through this works, we presented mobile edge cloud platform model and skeleton.

The rest of paper is organized as follows. In Sect. 2, we motivate the need for the mobile edge cloud platform based on related and previous works. In Sect. 3, we present the high-level design of the scalable mobile edge cloud platform and core components of the system. Finally, Sect. 4 discuss and conclude the paper.

2 Related and Previous Works

Habak et al. proposed FemtoClouds [6] system that provides a dynamic and self-configuring "multiple device mobile cloud system to scale the computation of cloudlet by coordinating multiple mobile devices. They consider how a collection of co-located devices can be orchestrated to provide a cloud service at the edge. Scenarios with co-located devices include, but are not limited to, passengers with mobile devices using public transit services, students in classrooms and groups of people sitting in a coffee shop. To this end, they propose the FemtoCloud system which provides a dynamic, self-configuring and multi-device mobile cloud out of a cluster of mobile devices. They present the FemtoCloud system architecture designed to enable multiple mobile devices to be configured into a coordinated cloud computing service despite churn in mobile device participation. They develop a prototype of FemtoCloud system and use it in addition to simulations to evaluate the performance of the system showing its efficiency and ability to leverage the available devices' compute capacity. This system contribute to a line of research on small, local and possibly private clouds.

Chen et al. proposed a distributed computational offloading model for Mobile Edge Computing [7]. They first study the multi-user computation offloading problem for mobile-edge cloud computing in a multi-channel wireless interference environment. They show that it is NP-hard to compute a centralized optimal solution, and hence adopt a game theoretic approach for achieving efficient computation offloading in a distributed manner. They then design a distributed computation offloading algorithm that can achieve a Nash equilibrium, derive the upper bound of the convergence time, and quantify its efficiency ratio over the centralized optimal solutions in terms of two important performance metrics.

Load balancing guarantees that all physical resources within cloud DC have a uniform workload. Existing load blanching schemes such as, Round Robin [8], Min-Min scheduling [9–11], Max-Min Algorithm [12], OpenStack Scheduler [13], Min-min Algorithm [14], and Improved Max-min [12], have considered static load balancing (single-resource) to co-locate VMs. All aforementioned schemes opted VM

placement scheduler that overlooks queuing user requests, and a fair-share algorithm enabled resources provisioning within data centers.

Nowadays, due to increasing demands of mobile users and hype in the popularity of mobile applications, incorporating high performance functionality features within the mobile application results in quick battery charge depletion due to high usage of mobile components. Accurate energy estimation helps application developers to investigate application power consumption behavior at earlier development stages. Consequently, developers redesign their application to optimize the utmost power consuming portion of application [15, 16]. Base cost energy of an instruction highly depends on the circuitry activated during instruction execution on mobile phone. Each instruction within ARM ISA triggers different component of the mobile phone. Moreover, for a single instruction, power consumption varies depending on type of op code, number of operands, operand type, and the mobile model used. For instance, base cost energy for LOAD instruction is higher than ADD instruction as memory operations are much expensive than simple ALU based operations [17].

3 A Scalable Mobile Edge Cloud Platform

The scalable mobile edge computing platform might extends available computing server resources which monitored and required from subscriber demands. And this model based on previous mobile edge cloud computing platform as following Fig. 1.

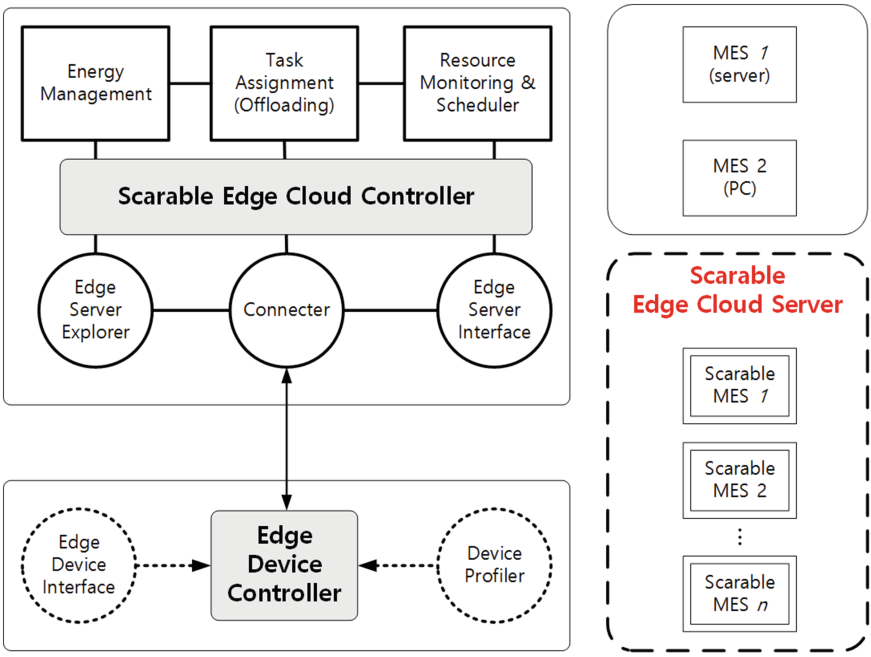


Fig. 1. An scalable mobile edge cloud platform and its workflow (bird-eye view).

The scalable mobile edge computing platform composed of the edge controller and edge device controller. The edge cloud controller comfortable connect and efficient resource manage between the edge device and edge server. The edge device controller manage the information which needed execute high efficient application and connect to the edge cloud. The resource monitor and scheduler maximize usage and access the computing resource between the edge servers regard to additional computing resources on the scalable edge cloud platform.

A task assignment offloader migrate a computing-intensive and energy-intensive modules using the static and dynamic code analysis technique for the execution speed enhancement and efficient energy consumption.

Finally, we designed the energy management modules that profiled the energy consumption information when mobile edge application execution time and motored the result of energy estimation for the minimization of energy consumption burden on the scalable edge cloud platform.

4 Conclusions

In this paper, we proposed and designed a scalable-edge cloud platform which an edge device subscribers could flexible extends cloud edge server to edge cloud and could application services specific to edge devices requirements in current static environments. And then, we verified the overall system completions and performances using the appropriate edge device application development. A scalable edge cloud platform composed of edge cloud controller and edge device controller. For the appropriate experimental edge device application, 360° panorama to plain view translation, moving object recognition, moving object tracking, we'll develop the 360° panorama image processing application that offloaded to rich resource computing server. Through this works, we presented mobile edge cloud platform model and skeleton.

Acknowledgments. This work was supported by the Engineering Research Center of Excellence Program of Korea Ministry of Science, ICT & Future Planning (MSIP)/National Research Foundation of Korea (NRF) (Grant NRF-2008-0062609), the IT R&D program of MSIP/KEIT. [K10047212, Development of homomorphic encryption supporting arithmetics on ciphertexts of size less than 1 kB and its applications], the Brain Korea 21 Plus Project in 2014, and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0024529).

References

1. Atzori, L., Iera, A., Morabito, G.: The internet of things: a survey. Comput. Netw. **54**(15), 2787–2805 (2010)
2. Satyanarayanan, M., Bahl, P., Caceres, R., Davies, N.: The case for vm-based cloudlets in mobile computing. IEEE Pervasive Comput. **8**(4), 14–23 (2009)
3. Satyanarayanan, M.: A brief history of cloud offload: a personal journey from odyssey through cyber foraging to cloudlets. ACM SIGMOBILE Mob. Comput. Commun. Rev. **18**(4), 19–23 (2015)

4. Ahmed, E., Akhunzada, A., Whaiduzzaman, M., Gani, A., Ab Hamid, S.H., Buyya, R.: Network-centric performance analysis of runtime application migration in mobile cloud computing. Simul. Model. Pract. Theor. **50**, 42–56 (2015)
5. Liu, J., Ahmed, E., Shiraz, M., Gani, A., Buyya, R., Qureshi, A.: Application partitioning algorithms in mobile cloud computing: taxonomy, review and future directions. J. Netw. Comput. Appl. **48**, 99–117 (2015)
6. Habak, K., Ammar, M., Harras, K., Zegura, E.: Femto clouds: leveraging mobile devices to provide cloud service at the edge. In: 2015 IEEE 8th International Conference on Cloud Computing (CLOUD), pp. 9–16, June 2015
7. Chen, X., Jiao, L., Li, W., Fu, X.: Efficient multi-user computation offloading for mobile-edge cloud computing. IEEE/ACM Trans. Networking (99) (2015)
8. Sidhu, A.K., Kinger, S.: Analysis of load balancing techniques in cloud computing. Int. J. Comput. Technol. **4**(2), 737–741 (2013)
9. Chen, H., Wang, F., Helian, N., Akanmu, G.: User-priority guided min-min scheduling algorithm for load balancing in cloud computing. In: 2013 National Conference on Parallel Computing Technologies (PARCOMPTECH), pp. 1–8. IEEE (2013)
10. Wang, S.-C., Yan, K.-Q., Liao, W.-P., Wang, S.-S.: Towards a load balancing in a three-level cloud computing network. In: 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, pp. 108–113. IEEE (2010)
11. Etminani, K., Naghibzadeh, M.: A min-min max-min selective algorithm for grid task scheduling. In: 3rd IEEE/IFIP International Conference in Central Asia on Internet, ICI 2007, pp. 1–7. IEEE (2007)
12. Elzeki, O., Reshad, M., Elsoud, M.: Improved max-min algorithm in cloud computing. Int. J. Comput. Appl. **50** (12), 9
13. Litvinski, O., Gherbi, A.: Openstack scheduler evaluation using design of experiment approach. In: 2013 IEEE 16th International Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing (ISORC), pp. 1–7. IEEE (2013)
14. Patel, G., Mehta, R., Bhoi, U.: Enhanced load balanced min-min algorithm for static meta task scheduling in cloud computing. Procedia Comput. Sci. **57**, 545–553 (2015)
15. Li, D., et al.: Calculating source line level energy information for android applications. In: Proceedings of the 2013 ACM International Symposium on Software Testing and Analysis (2013)
16. Shye, A., Scholbrock, B., Memik, G.: Into the wild: studying real user activity patterns to guide power optimizations for mobile architectures. In: Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture (2009)
17. Chang, N., Kim, K., Lee, H.G.: Cycle-accurate energy consumption measurement and analysis: case study of ARM7TDMI. In: Proceedings of the International Symposium on Low Power Electronics and Design, ISLPED 2000. IEEE (2000)

Network Anomaly Detection Based on Probabilistic Analysis

JinSoo Park, Dong Hag Choi, You-Boo Jeon, Se Dong Min, and Doo-Soon Park(✉)

SoonChunHyang University, RM U1202, 22 Soonchunhyangro, Shinchang-myeon, Asan-si, ChoongCheongNam-do, South Korea
{vtjinsoo, cdh0l9l, jeonyb, sedongmin, parkds}@sch.ac.kr

Abstract. In this paper, we provide a detection technology for a common type of network intrusion (traffic flood attack) using an anomaly data detection method based on probabilistic model analysis. Victim's computers under attack show various symptoms such as degradation of TCP throughput, increase of CPU usage, increase of RTT (Round Trip Time), frequent disconnection to the web sites, and etc. These symptoms can be used as components to comprise *k*-dimensional feature space of multivariate normal distribution where an anomaly detection method can be applied for the detection of the attack. These features are in general correlated one another. In other words, most of these symptoms are caused by the attack, so they are highly correlated. Thus we choose only a few of these features for the anomaly detection in multivariate normal distribution. We study this technology for those IoT networks prepared to provide u-health services in the future, where stable and consistent network connectivity is extremely important because the connectivity is highly related to the loss of human lives eventually.

Keywords: Anomaly detection · Network intrusion · Traffic flood · DDos attacks

1 Introduction

There are various techniques presented so far in order to detect or prevent the network attacks. And most of these techniques are based on the use of salient features such as changes in traffic volume, number of mismatched SYN packets, TCP throughput, round-trip time (RTT), CPU usage, and etc. to enhance the performance of the attack. Related studies are as follows. Authors in [1] present analytical studies on the flood attack on the web server to examine impacts of the flood attack in terms of TCP throughput, RTT and CPU usage. The TCP throughput degrades significantly when the system is under attack, the RTT is increasing, and the CPU usage is increasing as well. Mahalanobis distance is used to detect the normal and abnormal traffic, and the entropy of packet attributes (TCP flags and SYN/ACK packet rate per second) is used as features for the detection purpose [2]. Using Mahalanobis distance, traffic can be found suspicious when TCP packet distribution and SYN/ACK packet rate have sudden changes than expected. Similar techniques are employed in [3], which is based on the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_107

analysis of traffic entropy to provide Multilevel DDoS detection and prevention for DDoS attack, and multithreaded defense technology is suggested to process large amount of data and to reduce the packet loss for making the detection strategy doable.

A new feature to detect DDoS attack is proposed in [4] where the idea is to examine degree of unmatched SYN packets by observing bidirectional traffic flows, which may indicate a sign of attack. A flow–based intrusion detection method is proposed in order to detect various types of attacks and to reduce repetitious alarms caused by the same attack [5]. Unlike most existing anomaly detection methods, this research deals with groups of related packets together and examines similarities between the related groups to decide abnormality of the group of packets.

Most of these related studies emphasize the importance of chosen features, which can lead to the performance of detection systems. We focus more on the anomaly detection method rather than elaborating on choosing appropriate features. The idea is based on performing the anomaly data detection in multidimensional feature space assuming that data in the multidimensional feature space is highly correlated and can be modelled by a multivariate normal distribution somehow. Data anomalies can be determined by applying certain threshold for the multivariate normal distribution, so data under the threshold are suspected as anomalies.

This paper is organized as follows. Section 2 describes the anomaly detection strategy in multi-dimensional feature space suggested in this paper and discusses brief summary about the multivariate normal distribution. In addition, the expectation-maximization algorithm is explained briefly, which is used to estimate mean and variance of the multivariate normal distribution. Section 3 discusses simulation results of this research with several illustrations. Finally, we discuss conclusions of the study and future studies for later.

2 Detection of DDoS Attacks Based on Anomaly Detection Technique

2.1 Anomaly Detection

Most symptoms of DDoS attacks are at least approximately, correlated. For example, victim's computers show slow response to certain inputs, high CPU loads, frequent disconnection to web sites, low TCP throughputs, and etc. These symptoms are highly correlated, which are mostly occurred by the DDoS attacks. This correlated data can be represented in k-dimensional feature space, and it can be described by the multivariate normal distribution having density as follows [6].

$$f_X(x_1, \ldots, x_k) = \frac{1}{\sqrt{(2\pi)^k |\sum|}} exp\left(-\frac{1}{2}(x-\mu)^T \sum^{-1} (x-\mu)\right) \quad (1)$$

where x is a real k-dimensional column vector with k-dimensional mean vector μ and k X k covariance matrix $\sum$, and these components are represented by the following notation respectively:

$$\mathrm{x} \sim N\left(\mu, \sum\right)$$
$$\mu = [E[X_1],\ E[X_2], \ldots, E[X_k]]$$
$$\sum = [\mathbf{Cov}[X_i, X_j]],\ i = 1, 2, \ldots, k;\ j = 1, 2, \ldots, k$$

Detection of abnormal data is based on discovery of abnormal data in multivariate normal distribution density where data can be dealt with a set of correlated real-valued random variables. The idea begins by the notion that normal data are distributed near a mean value, however abnormal data can be located away from the mean in general. Thus if we find a proper threshold value (ε) to discriminate the normal and abnormal data in the multivariate normal distribution density, we can detect the anomaly of data. The corresponding algorithm can be summarized as follows.

1. Find the model $f_X(x_1, \ldots, x_k)$ on training data set $x^{(1)}, \ldots, x^{(k)}$ where k is the number of data
2. On a cross validation/test data x, predict the anomaly (y) of data as follows:

$$y = \begin{cases} 0 & if \quad f_X(x) < \varepsilon \quad (anomaly) \\ 1 & if \quad f_X(x) \geq \varepsilon \quad (normal) \end{cases}$$

2.2 Estimation of Mean and Variance by Expectation Maximization

Since the training data is assumed to be correlated in multidimensional feature space and can be modelled by multivariate normal distribution, then the problem can be considered as finding an appropriate Gaussian mixture model for single clustered data. Thus, we use the expectation-maximization (EM) algorithm to find the mean and the variance of the multivariate normal distribution. As one of the unsupervised learning algorithms, the EM algorithm is based on an iterative method to find maximum likelihood estimates of statistical models, where the unobserved latent variables comprise the statistical models [7]. The EM algorithm is comprised of an expectation (E) step and a maximization (M) step, and these two steps alternates until the latent variable estimation converges to a certain value. These two steps are summarized below.

Expectation step (E step): calculate the expected value of the log likelihood, with respected to the conditional distribution of Z given X under the current estimate of the parameters $\theta^{(t)}$:

$$Q(\theta|\theta^{(t)}) = \mathbf{E}_{\mathbf{Z}|\mathbf{X},\theta^{(t)}}[\mathbf{log}L(\theta; X, Z)]$$

Maximization step (M step): Find the parameters that maximizes this quantity:

$$\theta^{(t+1)} = \arg\max_{\theta} Q(\theta|\theta^{(t)})$$

3 Estimation Results

3.1 Training Data Generation

In [1], the symptoms of a UDP flood attack were analyzed in three different aspects: TCP throughput, average RTT, and CPU usage. When systems are under attack, the TCP throughput drops down significantly, and the average RTT as well as the CPU usage increases significantly. These symptoms can be used as features for the detection of network attack. More than two kinds of data can be used as features that will be represented as a random vector of a size k. However, this paper chooses only two different kinds of data (TCP throughput and CPU usage) as features because these features are caused by the attack and so they are highly correlated one another. Two highly correlated data can be fairly enough for the anomaly detection. And also it is convenient to see the detection results in 2-D space visually. Based on the description in [1], the TCP throughput gets dropped significantly during the attack while it is usually maintained at constant rate before the attack. Also the CPU usage increases significantly during the attack, but it is maintained at nearly constant percentage usually. When we generate these two data during the normal state, they must be correlated to each other as in [8]. Similarly, we generate these two data during the attack period based on normal distribution, but it has different mean and variance as shown in [1], where the TCP throughput and the CPU usage has significant difference during the normal and the attack period. For the normal TCP throughput data generation, we choose 94(Mbps) for a mean value and 1 for a variance respectively. And for the CPU usage data generation, we choose 40 for a mean value and 1 for a variance respectively. More details on this configuration are given in Table 1.

Table 1. Values assigned to generate normal type data

	Values	
TCP throughput	Mean	94
	Variance	1
CPU usage	Mean	40
	Variance	1
Correlation coefficient	ρ	−0.4
Number of data	N	100
EM	Number of clusters	1
	Initial points	random

Figure 1 shows several snapshots of estimated normal distribution in 2-D feature space with contours in different color corresponding to the same density value respectively. In other words, each contour has the same density value under a multivariate normal distribution. As shown in the figure, the TCP throughput and CPU usage data are negatively correlated and most data are located near the mean of two data. The expectation-maximization technique is implemented to estimate the mean and the variance of the density with configuration parameters given in Table 1. After this simulation, a threshold is chosen appropriately to accommodate all training data set.

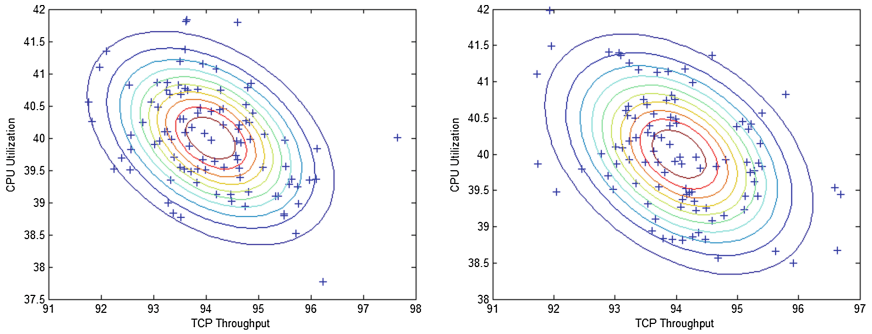


Fig. 1. Example of training data and corresponding EM results

3.2 Abnormal Data Detection Results

Abnormal data are generated based on the following configuration (in Table 2). Figure 2 shows both normal and abnormal data in the 2-D feature space and corresponding detection results where normal data are mostly located in lower right hand side and the abnormal data are mostly located in the upper left hand side (in red circle). With the huge difference of values during the normal and abnormal states, the

Table 2. Values assigned to generate abnormal type data

	Values	
TCP throughput	Mean	20
	Variance	1
CPU usage	Mean	200
	Variance	1
Correlation coefficient	ρ	−0.4
Number of data	N	100
Number of abnormal data	M	10

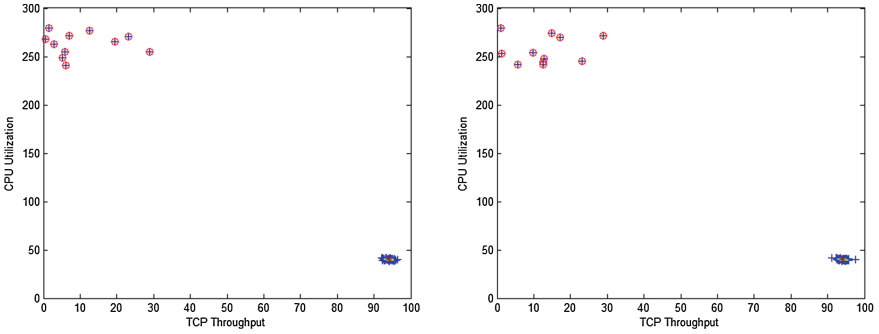


Fig. 2. Detection results on cross-validation/test data

corresponding density values for the abnormal data are significantly low compared to those of the normal data. So it is relatively easy to choose the threshold value to detect the anomalies because the density values for most of the abnormal data are nearly zero.

4 Conclusions and Future Studies

This paper presents an anomaly detection-based network intrusion detection strategy based on a probabilistic model. We use two anomaly symptoms (TCP throughput and CPU usage) as features in 2-D feature domain where we assume these features are highly correlated and can be modelled by the multivariate normal distribution with certain mean and variance. Then, the detection of the attack can be achieved by finding the anomalies of data under a chosen threshold value in the multivariate normal distribution density. We generate training data and test data respectively in order to verify the performance of the idea, which is found effective in analytical sense. For this study, we use the simulated training data to verify the idea of this paper, but we plan to apply this idea for real Internet trace file for experiments or to build an experimental test-bed for certain application such as IoT networks for u-health care services in our next studies.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP (Institute for Information & communications Technology Promotion)

References

1. Kolahi, S.S., Treseangrat, K., Sassafpour, B.: Analysis of UDP DDoS flood cyber attack and defense mechanisms on web server with Linux Ubuntu 13. In: 2015 International Conference on Communications, Signal Processing, and their Applications (ICCSPA), pp. 17–19, February 2015
2. Bayarjargal, D., Cho, G.: Detecting an anomalous traffic attack area based on entropy distribution and mahalanobis distance. Int. J. Secur. Appl. **8**(2), 87–94 (2014)
3. Rodgers, J.L., Nicewander, W.A.: Thirteen ways to look at the correlation coefficient. Am. Stat. **42**(1), 59–66 (1988)
4. Kejie, L., Dapeng, W., Fan, J., Todorovic, S., Nucci, A.: Robust and efficient detection of DDoS attacks for large-scale internet. Comput. Netw. **51**, 5036–5056 (2007)
5. Weon, I.-Y., Song, D.-H., Ko, S.-B., Lee, C.-H.: A multiple instance learning problem approach model to anomaly network intrusion detection. Int. J. Inf. Process. Syst. **1**(1), 14–21 (2005)
6. https://en.wikipedia.org/wiki/Multivariate_normal_distribution
7. https://en.wikipedia.org/wiki/Expectation-maximization_algorithm
8. http://kr.mathworks.com/matlabcentral/answers/231480-how-to-generate-random-numbers-correlated-to-a-given-dataset-in-matlab

Hedonic Model Study for Retargeting Advertising Based on Space-Centered Internet of Things

Bo-Ram Kim[1], Man-Soo Chung[1], and Yong-Ik Yoon[2(✉)]

[1] Department of Public Relations Advertising,
Sookmyung Women's University, Seoul, Korea
{Rang9l45, mchung}@sm.ac.kr
[2] Department of IT Engineering,
Sookmyung Women's University, Seoul, Korea
yiyoon@sm.ac.kr

Abstract. This paper is focused on hedonic model study for retargeting advertising Based Internet of Things using useful information. Many research related to the existing Internet of things, relatively not many study for effective advertising model based Internet of Things. So, this paper is designed more information, fun, interactive advertising model based on Internet of Things. Therefore, result of this paper show that implication to produce advertising based on Internet of Things provides a practical guide.

Keywords: Internet of things · Retargeting advertising

1 Introduction

Today we live in a rapid change of new technology developed every day, and along the innovative change of the Internet and mobile that we've experienced in the past, now we are facing another change. Past advertising tried to deliver the message to the consumer in one-way, but now many company prefers to do the retargeting advertising, which prefers to select only the appropriate consumers that have more purchase intention on a particular item and resend the advertising message to them. This way of retargeting advertising, as the use of Internet and Mobile has been increased, utilizes the individual consumer's information based on cookies to use their online trace to execute advertising into desktops, lab tops, smart phones, tablet PC, and so on that each individuals access. However, advertising based on the current Internet of Things is only following the simple message formation that supplies simple information only, the evolved smart retargeting advertising that goes beyond this, as the further IoT technology becomes pretty close to our lives. Hence, this paper we suggests Hedonic Model for retargeting advertising based Internet of Things, which is a new kind of advertising as applying the concept of Hedonic Model that enables purchase behavior by being useful, fun and interesting one instead of just a simple advertising message.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_108

2 Research Background

Hedonic model for retargeting advertising, which is based on the space-centered IoT, that our paper suggested is a model that resends the advertising of the product that got exposed and clicked by consumers based on their online information, but not purchased, on the Internet of Things with additional fun and interesting elements together. Especially when retargeting advertising based on the space-centered IoT is active, it creates a new platform that could expose advertisement to consumers much more. As consumers see appropriate advertisement on where they stay, this is the implication of a potential alternative that decreases fatigability from constant exposure to unwanted advertisement.

2.1 Internet of Things and Advertising

Internet of Things is a term that Kevin Ashton first used, which refers to the technology that connects things from our surroundings to the Internet that enables to offer additional worth to consumers. According to Gartner during market research phase in America, it is assumed that 100 Billion of things and device will connect to Internet in 2020. Today new platforms are constantly formed into online area and advertising industry is facing new changes every day as the increased use of smart phone. In particular, if IoT, the new technology becomes active, it will connect not only just to the prior PC, but also to all the spaces as well as cars or home appliances, and another new digital innovation will come after the Internet and mobile innovation. It gives a new platform of advertisement to advertising industry as well as our real-life setting, and many changes are expected to utilize it (Fig. 1).

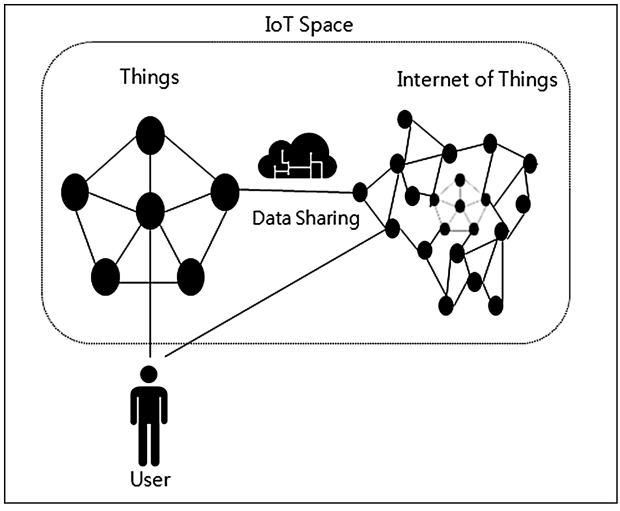


Fig. 1. Internet of things based IoT space

Due to the change in this new setting, consumers seem to be changing in such as their new shopping behaviors or collecting information aspects. Today retargeting advertising, which is a way to offer a necessary advertising to the selected consumers that are in search of or interested in products of particular company by precise targeting using consumers' information, begins to be active by global companies such as Google, and it is taking center stage of the most effective and smartest advertising tool. Figure 2 shows the Feature of retargeting advertising than previous one way advertising that afford the space-centered selected advertising from Internet of Things to the particular individuals. In prior setting, one-way advertisement was preceded to many consumers, and their individual data are not structure and connected to each other since it is not systemized data. But as collected pieces of data onto particular individual are systemized and more meaningful information is created, it enables to make customized advertisement to each individual.

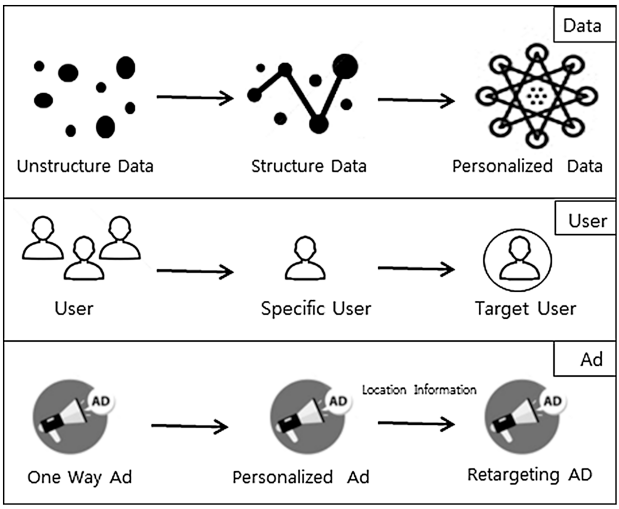


Fig. 2. Retargeting advertising feature

Also it's a sensor that is connected to all things IoT that it enables appropriate customized advertising to particular space as the information is finely connected. If this kind of retargeting advertising formation is used in advertising that is based on space-centered IoT in the future, we expect that it will attract the consumers that are appropriate in more various spaces or platforms effectively. Also, because the retargeting advertising, which is exposed constantly to consumers, could make them tired, the additional informative, fun and elements that effectuate their participation of Hedonic model will help the current retargeting advertising to be evolved again in a smarter way.

2.2 Retargeting Advertising Based on Internet of Things

Hedonic model for retargeting advertising on Internet of Things should offer consumers the advertising as useful information, and also make them interested in it despite of constant exposure, so that they eventually purchase the product instead of getting tired of it. The prior retargeting advertising model resends the customized advertisement around user's information. It should be noted that Hedonic model for retargeting advertising based IoT, unlike the prior one, offers advertisement considering elements such as interest, fun, information, or interactivity, which is creative representation technique of advertising. Its schematization is shown on Fig. 3.

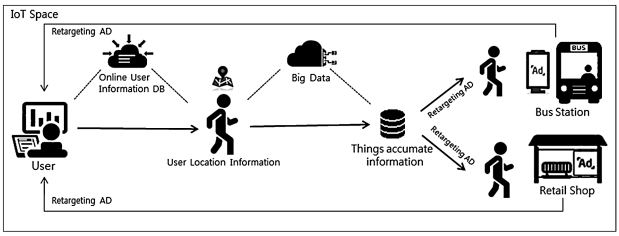


Fig. 3. Retargeting advertising system based IoT

In other words, Personal indistinguishable information of what and where users searched or clicked or visited in order to find particular information online will be collected as DB. If users move on to another space at this, the collected information and space information are analyzed and combined as big data onto Internet of Things, and finally shown in the selected appropriate space as advertisement again.

So more useful and fun advertisement should be offered to consumers, as more interesting and smarter informative advertisement at this point where Hedonic model for customer-driven retargeting advertising based on IoT instead of online has not been activated yet.

3 Strategy Method

3.1 Suggestion to the Model for Smart Retargeting Advertising

This paper suggests an advertising method that could effectively apply smart retargeting advertising that utilizes space centered IoT based on the elements presented above. This model reduces unnecessary advertisement of information while our surrounding things connecting to the Internet around the Internet of Things, and offers the selected information that was optimally customized to each individual. In other words, as the advertising is useful and has a fun and interesting way, this model prevents advertisement avoidance or irritation and induces purchase, and builds trust and satisfaction. First of all, the use pattern of users such as searching and clicking or visiting websites in order to look at particular products that they are interested or wanting to

purchase, will be collected as DB. Based on the information on users collected this way, it selects just the ones that users have directly taken a look at or shown interest, and reoffers them to induce purchase. The important thing at this point is that if a user searched online for a particular grocery and left without purchasing, it sends the advertisement through the user's refrigerator, dining table, or kitchen electronic appliances. That is, the concept of retargeting advertising is offering advertising with finely selecting the information that is only necessary to the users, and Hedonic model for smart retargeting advertising adds more fun and interesting elements and materialized advertising in space based on the Internet of Things, as shown in Fig. 4.

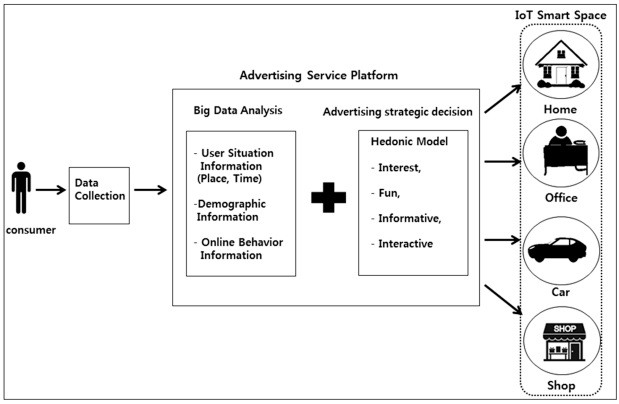


Fig. 4. Retargeting advertising system based Hedonic model

Adopting this kind of advertising will enable finely targeting considering consumers' life style, propensity and so on, and it will be able to be an alternative to reduce advertisement avoidance as arousing interest instead of a constant one-way advertisement. Also when Hedonic model of smart retargeting advertising is activated in the future, it will enable more effective and smarter advertising as selecting the most efficient medium. The future study about space-centered advertising based Internet of Things should apply psychological variables of particular consumers considering more various side of them as more of an in-depth study. Also due to the space-centered properties of IoT, the future study about advertising effects on particular place, time, or situation is also expected.

4 Conclusion

In this paper, we deal with Hedonic model for retargeting advertising based on space-centered IoT and products that consumers have directly visited or clicked but not purchased yet, in the Internet of Thing setting based on user's online information, with additional fun and more interesting elements for more useful advertising in the proper place, as shown in Fig. 5.

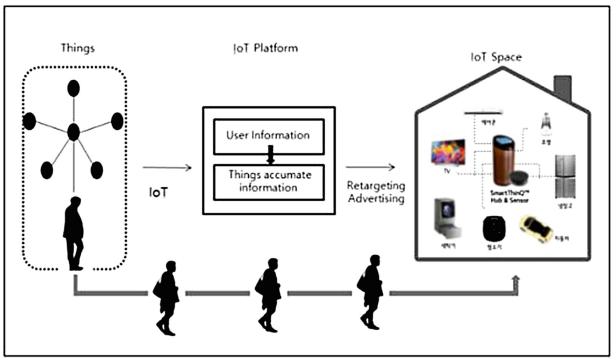


Fig. 5. Model of retargeting advertising in IoT space

From the point of consumers, we propose this advertising model is assumed that they would feel less irritated than existing one-way advertising, since it is a creative advertisement, which they've been interested before. Especially if the advertising based on the Internet of Things is activated, that means a new platform that could deliver the advertising is created, and consumers will be exposed to advertising much more than now. This paper has the implication of a potential alternative that decreases fatigability from constant exposure to unwanted advertisement. A advertising strategy based on IoT has been suggested in this paper that Hedonic model for smart retargeting advertising that applied the Internet of Things technique, but in the future study, it will proceed to select the appropriate consumer on Hedonic model for smart retargeting advertising, and plan to conduct the practical study to measure the effect after that.

Acknowledgement. This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0126-16-1041, Auto-Generated Media Service Technologies based on Semantic Relationship of Contents for Self-Growth Social Broadcasting).

References

1. KISA: Market Policy Trend Analysis of Internet of Things (IoT). Internet & Security Issue (2014)
2. Yoon, Y.I., Lee, S.J.: A study of advertising model based on hybrid user context in smart space. J. Korea Soc. Comput. Inf. **17**(2), 188–195 (2012)
3. Retargeter: What is Retargeting and How Does It Work? (2015). https://retargeter.com/what-is-retargeting-and-how-does-it-work
4. Hong, J.H., Cho, S.H.: Study of relation between consumers' advertisement attitude and need for cognition for IoT-implemented advertisement. Korea Digital Contents Soc. **16**(1), 165–175 (2015)
5. Etnews: Growth of M2M communication industry (2010)
6. Pyun, S.J., Jin, H.H.: The internet of things the huge connection surpass over cloud and big data future. Window Publishing Co, Chicago (2014)

7. Kim, J.D.: The internet of things era: status and prospects. Int. Trade Bus. Inst. **106**, 1–213 (2015)
8. Um, J.H., Park, J.K.: A study on the interaction paradigm shift in environment of internet of things focusing on smart device. Korea Sci. Art Forum., 471–487 (2015)

A New Automated Cell Counting Program by Using Hough Transform-Based Double Edge

Jae Sung Choi[1], Moon Jong Choi[2], Jung-Min Lee[1], and Hyun Lee[1(✉)]

[1] Department of Computer Science and Engineering, Sunmoon-ro-221, Tangjeong-myeon, Asan-si, ChungCheongnam-Do 31460, South Korea
{jschoi,mahyun9l}@sunmoon.ac.kr, wjaalslll@naver.com
[2] Department of DGMIF, cheombok-ro 80, Dong-gu, Daegu 41061, South Korea
mjchoi0@gmail.com

Abstract. A suitable amount of cells in a range is necessary in order to conduct the experiment. In addition, various methods are being performed to counter the number of cells. However, there are still some problems. We propose a new automated cell counting program by using Hough Transform-based double edge. The proposed algorithm can distinguish between dead cells and living cells automatically. Finally, we will show the improvement of our work by reducing the range of error rates.

Keywords: Automated cell counter · Hough transform · Double edge · Image processing · Smoothing

1 Introduction

In general, a suitable amount of cells in a range is necessary in order to conduct the experiment such like a toxicity testing, a stem cell, and an animal cell culture. In addition, various methods such like an electrical impedance, an image processing, and a selective photothermolysis [4] are being performed to counter the number of cells. In particular, there is a representative cell counter program by using image processing such like 'automated cell counter R1' [1] and 'LUNA^{TM} automated cell counter' [2].

However, there are still some problems for automated cell counting programs. For example, (1) 'Coulter Counter' [3] that utilizes electrical impedance calculates the agglomerated cells into single cells. In addition, it can't distinguish between dead cells and living cells automatically. To overcome the disadvantage of 'Coulter Counter', image analysis counters are used for automated cell counting program. One of methods is a selective photothermolysis and the other is the image processing such like 'R1' and 'LUNA'. For instance, (2) a selective photothermolysis can reduce the range of error rates because the person checks the image directly with the naked eye and can adjust the number of cells. However, it can only counter the dead cells or the living cells separately as shown in Fig. 1. Moreover, (3) 'R1' and 'LUNA' can reduce the range of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_109

Fig. 1. Manual cells counting and existing method for cell counting using image analysis

error rates and can distinguish between dead cells and living cells. They can also save the data using the USB device. However, the price of 'R1' and 'LUNA' device is too high and then the cost efficiency is too low. The general laboratory researchers can have problem to use the system personally.

Therefore, in this paper, we propose a new automated cell counting program by using Hough Transform [5] -based double edge. The proposed algorithm can reduce the range of error rates and can distinguish between dead cells and living cells. It can also improve the cost efficiency compare to the price of 'R1' and 'LUNA' device. Finally, we develop the automated cells counting program based on the proposed method and we analyze the proposed method in this work.

The rest of this paper is organized as follows. In Sect. 2, we introduce the proposed algorithm and system process. In Sect. 3, we analyze the experimental results and then we discuss our conclusions and future work in Sect. 4.

2 The Proposed Algorithm

2.1 System Process

In general, Hough Transform is used to find the features such like a straight line, curves, and circles in the image. Additionally, edge detection in the image processing is used to get the selected range from the image. In this work, we combine two methods and then we try to get circles from the image using Hough Transform-based double edge algorithm. For example, first, we find the cells from the image by using edge detection processing. Depending on the contour information of the cells, we can find the living cells from the image. Second, we dye the detected cells to distinguish between dead cells and living cells. The dead cells have only one circle. The living cells have two circles in the image. Third, we find the dead cells from the all cells by operating double edge detection processing. The dead cells don't have any two circles in the image. Finally, we calculate the dead cells and the living cells then we store the information to the system.

2.2 Hough Transform-Based Double Edge Algorithm

As shown in Fig. 2, the proposed Hough Transform-based double edge algorithm is consists of five steps. First, the noise reduction processing is performed. If the noise reduction processing isn't performed, it is difficult to detect the circles from the image. Second, the smoothing [6] processing is performed to mitigate the damage of image. In this work, we utilize Gaussian filtering techniques [7] because of Gaussian filter by convolution with a Gaussian kernel in each of the pixels generates a resulting image. Third, we perform canny edge [8] processing to obtain binary image. To make this, we convert the image to grayscale image then the grayscale image convert to binary image. After canny edge processing is performed, fourth, we select the circles from the image by using Hough Transform. At this time, we need to set the range of the distance between the circle and the circle. In particular, we need to set the size of the minimum radius of the circle and the size of the maximum radius of the circle in order to reduce the error rates. After the circle detection is completed, finally, we set a new threshold value to convert the image then to perform double edge detection and image processing as shown in Fig. 3.

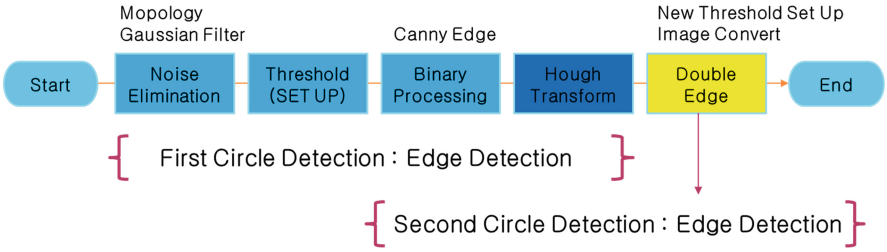


Fig. 2. System process (Five steps) of Hough Transform-based double edge

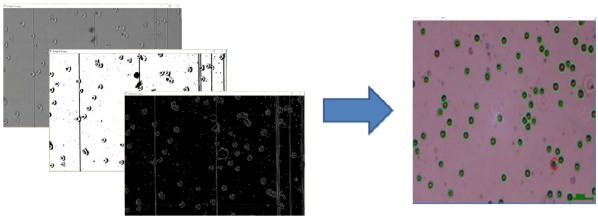


Fig. 3. An example of Hough Transform-based double edge algorithm

3 Experiment and Analysis

We developed a cell counting program that composed of OpenCV with MFC for making experiments. In addition, the cells were tested by taking photos and videos into the program. We assume that the living cells are expressed by green circle and the dead

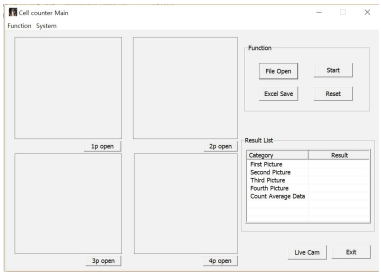
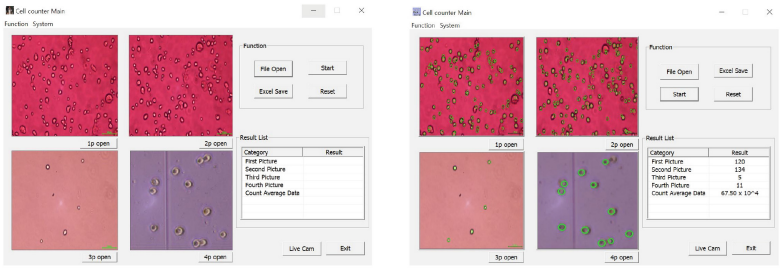


Fig. 4. An example of the developed cell counting program (Default)

cells are expressed by red circle. As shown in Fig. 4, we can put four images on the left and it exhibits a total calculated value and average on the right.

As the experiment results, Figs. 5(a) and 5(b) show an example. In the figure, the green circles are living cells and the red circles are dead cells. In this case, the error rates are $\pm$ 5 %.



(a) Image matching (b) Circle Detection

Fig. 5. An example of the developed cell counting program (Success). (a) Image matching (b) Circle Detection

In addition, as shown in Fig. 6(a) and 6(b), real-tiem video source detection takes palce over the original binary image. Figure 6(a) shows an example of the live video screen and Fig. 6(b) shows an example of the circle detection in real-time image.

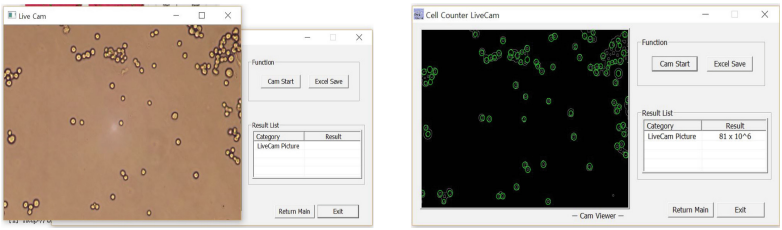


Fig. 6. An example of the developed cell counting program (Success)

4 Conclusion

In this paper, we proposed a new automated cell counting program by using Hough Transform-based double edge. In particular, we developed the proposed cell counting program based on MFC with OpenCV to make an experiment. The proposed method distinguished between dead cells and living cells. Also, it reduced the range of the error rates up to ± 5 %. Moreover, it improved the cost efficiency compared to 'R1' and 'LUNA'.

However, it consists of simple cell counting mechanism compared to 'R1' and 'LUNA'. In addition, the circle detection is performed by using the binary image in real-time case. Thus, we will upgrade current version of the developed cell counting program in order to add more function to the system. Furthermore, we will modify the system to catch the circle detection based on the real image in real-time case.

Acknowledgments. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013R1A1A1075980).

References

1. R1. http://www.olympus-lifescience.com/ko/cell_culture_solution/r1/
2. LUNA. http://logosbio.com/cell_counters/luna/features.php
3. Graham, M.D.: The coulter principle: foundation of an industry. J. Assoc. Lab. Autom. **8**(6), 72–81 (2003)
4. Oh, S., Park, J.: Physical Properties of Living Cells, pp. 1–9, February 2014. DOI:10.3938/PhiT.23.001
5. Hassanein, A.S., Mohammad, S., Sameer, M., Ragab, M.E.: A survey on hough transform, theory, techniques and applications. IJCSI **12**(1), 1–18 (2015)
6. Smoothing. https://terpconnect.umd.edu/∼toh/spectrum/Smoothing.html
7. Gaussian Filter. https://en.wikipedia.org/wiki/Gaussian_filter
8. Canny Edge Detector. http://www.cse.iitd.ernet.in/∼pkalra/csl783/canny.pdf

An Approach for Interworking Heterogeneous Networks with DTN and IP Routing in Space Internet

Euiri An, Kyungrak Lee, Jaewon Lee, and Inwhee Joe(✉)

Mobile and Network Convergence Laboratory, Computer Science Department, 222 Wangsimni-ro, Seongdong-gu, Seoul 04763, South Korea
{reegoon, esilote82, kallontz, iwjoe}@hanyang.ac.kr

Abstract. In a space mission, there are various types of nodes which can adopt individual communication protocols for aircraft on Delay Tolerant Network (DTN) and terrestrial control center on TCP/IP. For efficient routing between heterogeneous networks, we propose an approach for interworking heterogeneous networks with DTN and IP routing in space Internet. We design it by configuring Interplanetary Overlay Network (ION), based on IP forwarding rule and gateway. Also, we design scenarios to evaluate its performance. The experimental results verify the proposed protocol is suitable for interworking heterogeneous networks in space Internet.

Keywords: DTN · ION · Heterogeneous networks · IP forwarding · Interworking network

1 Introduction

There are several characteristics in space Internet such as long propagation delay, intermittent disconnection and node's forcible moving path due to orbit of celestial bodies. To solve that problems, Delay Tolerant Network (DTN) has been suggested and relevant researches are ongoing [1, 2].

A network topology which will be deployed in a space mission is comprised of terrestrial control center, orbiter and probe. They are based on different networks. Therefore, it's difficult to apply an existing routing protocol that has been using on terrestrial Internet. This interworking issue is one of the first problem in efficient aspects with routing. But, there's no research on that interworking problem.

In this paper, we design a routing protocol for interworking with DTN in space Internet and IP network in terrestrial Internet. For this, by applying IP forwarding, interworking algorithm is designed which interworks with IP node and DTN node. It enables that heterogeneous nodes can relay their transmission. Then we design several test scenarios to evaluate its performance. With this, we propose DTN routing protocol supporting various network structures which will be deployed in a space mission.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_110

2 Related Works

Due to some characteristics in space Internet such as long propagation delay and intermittent disconnection, new network system is suggested which is DTN [1, 2]. JPL developed Interplanetary Overlay Network (ION) that implements DTN environment with Bundle Protocol (BP). It is implemented to apply in a space mission, so it supports space simulations on the ground [3].

Bundle Protocol (BP) is a core transmission protocol with Store-and-forward transmission. It doesn't require a full connection from source node to destination node. Relay node stores data and transmits it one by one hop. Against the disconnection, Licklider Transmission Protocol (LTP) supports transmission on various paths and retransmission. For this, it provides reliability [2].

Due to these characteristics, an appropriate routing protocol is required in space Internet. There are several routing protocols based on DTN such as Epidemic, Prophet and Spray and wait.

First, Epidemic routing copies messages when node contacts any other nodes unconditionally (flooding). It shows inadequate performance on a size of buffer and network traffic aspects [4]. Prophet routing uses delivery probability by calculating contact information with nodes. Depending on a priority, it shows high delivery probability [5]. Spray and Wait limits message copying, so it no longer copies message when it comes in wait phase. This shows outstanding performance by solving a critical problem in Epidemic routing [6].

Also, in [7], it calculates a priority with properties of heterogeneous node groups. With this, it proposes routing protocol which delivers a message. In addition, JPL suggested CGR. It synchronizes contact information ahead and it can avoid unnecessary changes on contact information. With this, each node can process Bundle forwarding based on contact plans [8].

3 Architecture for Interworking Heterogeneous Networks

3.1 Interworking Protocol

In this paper, we propose DTN routing protocol which enables interworking heterogeneous networks. First, we interwork with DTN network and terrestrial IP network. Also type and number of relay node should be considered in a multi-hop transmission. With this routing algorithm, these various networks can be interworked.

3.2 Routing Algorithm

In ION, node number is classified by using Endpoint ID (EID). We can configure transmission plans with EID ahead. There are steps configuring transmission path in a multi-hop transmission.

A. Configuration for ION Multi-hop Transmission

Above all, transmission plan rule should be configured on 'ipnrc' configuration file in ION. Here comes a form of plan rule.

$ a plan N ltp/M

'a' is a command adding a plan rule. 'plan N ltp/M' means that file is transmitted by LTP Outduct M when destination node is N. In this way, we can configure a multi-hop transmission between ION nodes.

B. Configuration for IP Forwarding Rule

In sequence, for interworking IP node, IP forwarding rule should be configured in a IP node. In this paper, we use Iptables which configures firewall based on Linux. Here comes a command configuring forwarding rule with Iptables.

$ Iptables –A PREROUTING –t nat –p udp –d RelayIP –dport RelayPort –j DNAT –to DestinationIP:DestinationPort

In this, RelayIP and RelayPort mean IP address and port number of relay node. In other words, these are incoming packet's destination parameters. The command, Iptables, changes them to DestinationIP and DestinationPort. Then it retransmits the packet.

C. Gateway Design for Interworking Heterogeneous Networks

Interworking nodes based on heterogeneous networks is required. In these network structure, there is a node which is comprised with more than two network interfaces. In this case, other network's gateway address should be added in a routing table. Here comes a command adding routing table based on Linux.

$ route add –net B_DefaultGateway netmask 255.255.255.0 dev InterfaceNumber

In a routing table of A, default gateway address of B is added with its interface number. With these processes such as ION multi-hop configuration, IP forwarding configuration and interworking gateway design, we can interwork various network structures. Using these algorithms, we design several test scenarios and evaluate their performance.

4 Experimental Results

4.1 Test Environment

In Table 1, there are environment properties of our simulation. In IP node (gateway), we turn firewall off for configuring IP forwarding. We compare performance with TCP and LTP. Each test is performed 10 times and we get average value by Treammean (calculate the average without maximum and minimum value).

4.2 Test Scenario and Results

A. Test Scenario 1: Data rate depends on types of relay node on a fixed protocol (LTP)

In Fig. 1, there is a testbed scenario 1.

Table 1. Test environment

Parameter	Contents
OS	Ubuntu 14.04 LTS
ION version	3.4.0b
Firewall	Gateway Node: Disabled
Number of nodes	3 ~ 8
Number of tests	10 each, Trimmean Average
Network interface	WLAN (802.11n)
Protocol	TCP, LTP
Size of message	1 MB ~ 10 MB

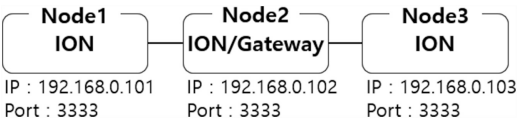


Fig. 1. Testbed scenario 1

In a relay node, other conditions are same but only different factor is that ION runs or not. In scenario 1–1, using plan rule of ION, we perform multi-hop transmission. In scenario 1–2, Multi-hop transmission is performed with proposed IP forwarding rule. In Fig. 2, there is a result of scenario 1. Even with IP relay node, it shows a bit better performance than ION relay node. It means that packet is forwarded on IP layer without reaching at application layer. As a result, with proposed algorithm, either IP node or ION node is suitable for a relay node. Also both types of node show similar throughput.

B. Test Scenario 2: data rate depends on types of protocol with a fixed IP relay node (Gateway)

In scenario 2, we configure different base protocol in ION with TCP and LTP. Other conditions are same with scenario 1–2.

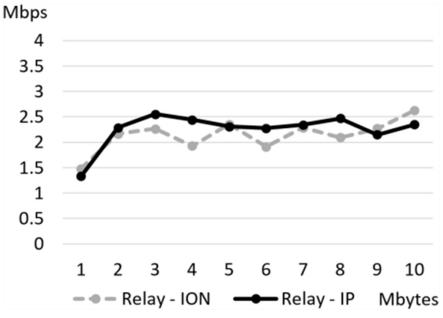


Fig. 2. Data rate as a function of the size of messages

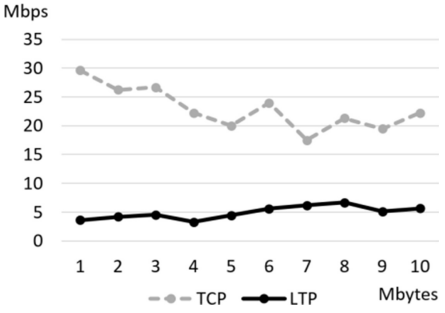


Fig. 3. Data rate as a function of the size of messages

In Fig. 3, it shows result of scenario 2. There are notable differences of performance between TCP and LTP. On TCP, it shows better throughput in multi-hop transmission.

C. Test Scenario 3: Propagation delay depends on number of IP relay nodes (Gateway) on a fixed protocol (LTP)

In scenario 3, we perform several multi-hop transmissions with various number of hops. As in Fig. 4, source and destination nodes run ION with BP over LTP over UDP, and relay nodes configure IP forwarding rule. The number of hops are 3 to 8 (Gateway: 1 ~ 6). Totally we perform six cases tests.

Fig. 4. Testbed scenario 3

In Fig. 5, there is a result of scenario 3. While increasing the number of relay nodes from 1 to 6, the propagation delay is also risen up as the number of nodes increase.

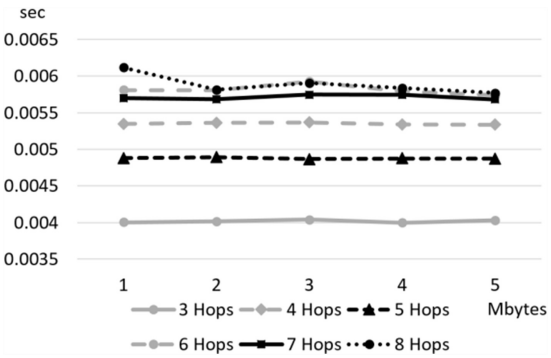


Fig. 5. Propagation delay as a function of the size of messages

5 Conclusion

In this paper, we design and evaluate interworking protocol for efficient routing operation. We configure IP forwarding rule and verify that IP node is suitable as a relay node in a transmission between ION nodes according to test results. Also we compare performances based on TCP and LTP protocols and evaluate performances according to the number of relay nodes in multi-hop transmission. With this, we confirm that the proposed interworking protocol is suitable for heterogeneous networks with DTN and IP routing in space Internet.

Acknowledgement. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (Ministry of Science, ICT & Future Planning) (No. 2016R1A2B4013118).

References

1. Delay Tolerant Networking Research Group. http://dtnrg.org
2. Fowjiya, S., Udhayachandrika, A., Kathirvel, A.: Architectural overview of delay tolerant network. In: IJESRT (2013)
3. Burleigh, S.: Interplanetary overlay network: an implementation of the DTN bundle protocol. In: IEEE Consumer Communications and Networking Conference, pp. 222–226 (2007)
4. Vahdat, A., Becker, D.: Epidemic routing for partially connected Ad-hoc networks. In: Technical Report CS-2000-06. Duke University (2000)
5. Lindgren, A., Doria, A., Davies, E., Grasic, S.: Probabilistic routing protocol for intermittently connected networks. In: IETF (2012)
6. Tambe, S.D., Chawan, P.M.: Analytical study of spray and wait routing protocol in delay tolerant network. In: IJATES (2014)
7. Spyropoulos, T., Turletti, T., Obraczka, K.: Routing in delay tolerant networks comprising heterogeneous node populations. In: IEEE Mobile Computing (2008)
8. Birrane, E., Burleigh, S., Kasch, N.: Analysis of the contact graph routing algorithm: bounding interplanetary paths. Acta Astronaut. **75**, 108–119 (2012). Elsevier

Implementation of Recommender System Based on Personalized Search Using Intimacy in SNS

Jeong-Dong Kim, Bongjae Kim(✉), and Jeong-Ho Park

Department of Computer Science Engineering,
Sun Moon University, Asan, Republic of Korea
{kjd4u, bjkim, jhpark}@sunmoon.ac.kr

Abstract. Recently, a search system has been a trend of personalization such as recommendation systems and social searches. Because, each users receive different results for the same queries by using user preference and interesting. Specially, a social relation is a most important factor of search system, and therefore, many recommender system using have been proposed. However, existing recommender systems typically return a set of search results based on a user's query without considering user interests and preference. Therefore, the identical query from each user will generate the same set of results displayed in the same way for all users. To overcome this restriction, this paper proposes a recommender system based on personalized search using intimacy in SNS and describe a prototype of our recommender system.

Keywords: Recommender system, personalized search · Polarity analysis · Intimacy · Social network services

1 Introduction

Social Network Service (SNS) is a platform to build social networks or social relations among users that is able to generate and share a large volume of information by real time [1]. Most SNS are web based and provide means for users to interact over the internet. As SNS continue to increase, more contents are created by users. Therefore, the SNS which has a variety of characteristics is generated through voluntary participation of users, which is also called 'Big Social Data' [2]. As the importance of web search has been recognized, extensive studies have been conducted actively to analyze the data in a SNS [3]. A SNS-based search research on finding new knowledge from a large amount of information that can identify not only contents registered in the web but also contents of the author as well as the friends of the user [4].

Most people use search system that can be used an integral part of our daily lives. Nowadays, there has been a trend of personalization search system such as recommendation systems and social searches. Because, each users receive different results for the same queries by using user preference and/or interesting. Existing recommender systems typically return a set of search results based on a user's query without considering user interests or preference [5]. Therefore, the identical query from each user will generate the same set of results displayed in the same way for all users.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_111

On the social network, the relations among people are connected by link, so we can know information of others who are both directly and indirectly related to themselves and use their data in personalized recommendation system. When social network data is applied to personalized recommendation system, the most strong point is that recommendation using the data of people who are directly related to become possible. If this recommendation system is set up, the accuracy would be much higher than existing system not using people relationship [6]. Because people tend to trust opinion of acquaintance who are related to, using people relationship in personalized recommendation system brings about high accuracy of recommendation. Accordingly, in this paper we propose an implementation of recommender system based on personalized search using intimacy in SNS.

2 Related Words

Social network data has various characteristic such as real-time, people relation and big social data. Because variety of social network data can be used to enhance recommendation, how to use social network information has been extensively studied for personalized recommendation system.

The majority of the existing work on recommender systems are focused at the contents based, fuzzy, semantic web, and web server logs. In this paper, we describe some of the recommender systems existing in Table 1. Table 1 presents different recommender systems used, domain focused and approach.

Table 1. Different recommender systems.

Recommender systems	Domain	Approach
YourNews [7]	Newsgroups	Content based
PRemiSE [8]	News	Social Experts
FTCP-RS [9]	Telecom	Fuzzy
PerHSS [10]	Hotel	Semantic web
WebPUM [11]	Web	Web server logs

3 Our Approach (Personalized Search System)

In this section, we describe the system overview for the proposed Personalized Recommender System (PRS) and measurement of user's intimacy using Twitter contents. We describes characteristics of the Twitter and proposed a methods for polarity and intimacy analysis of contents. The proposed PRS is based on user-based collaborative filtering approach. Differently from previous system, proposed recommender system considers intimacy calculating similarity between users. Existing collaborative filtering calculates similarity through rating of unspecified individual users for item that approach is less accuracy because of using unrelated people information.

The PRS consists of the modules *Crawling*, *Personalized Recommendation*. Figure 1 shows a system overview of the structure for proposed system.

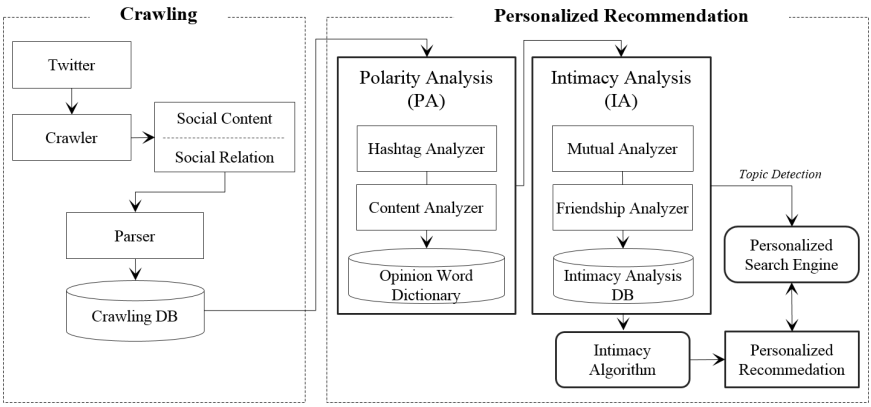


Fig. 1. Overview for PRS.

Crawling. The Twitter contents consists '*Social Content*' and '*Social Relation*'. The *Social Content* was composed with combinations such as content, written time, RTcount, and so on. Also, the *Social Relation* indicates among users/friends relations. The Twitter dataset are collected using the streaming API, parsing and then stored in *Crawling DB*.

Personalized Recommendation. In this module, various preprocessing steps are performed. The *Polarity Analysis* is generated by *Hashtag Analyzer*, and *Content Analyzer*. We development of our own *Opinion Word Dictionary (OWD)* expand the SentiWordNet [12] for analysis of Korean contents in Twitter dataset. Also we are analyzed intimacy of similarity of friends and friend relation levels by using *Mutual Analyzer* and *Friendship Analyzer*.

For *Polarity Analysis*, we use SentiWordNet, a lexical resource for opinion mining that is associated with three sentiment scores in each WordNet synset [13]. The SentiWordNet method defines L as the union of three seeds (i.e., training) sets, L_p and L_n of known *Positive*, and *Negative* synsets, respectively. Each ternary classifier is generated using the semi-supervised method presented. A semi-supervised method is a learning process whereby only a small subset, $L \subset T_r$ of the training data T_r have been manually labelled. Initially, the training data in $U = T_r - L$ are unlabeled. The process itself labels them, automatically, using L as input. L_p and L_n are two small sets, which we defined by manually selecting the intended synsets.

For the OWD, the polarity of Twitter contents is analyzed. We develop our own OWD by exploiting SentiWordNet. Using SentiWordNet, we find representative Korean vocabulary representing *Positive* and *Negative*, then add and modify the Twitter dataset; content with ambiguous opinion are classified as *Neutral*. Thus, Twitter contents are classified into *Positive*, *Negative*, and *Neutral*. Table 2 shows a part of the OWD for Korean polarity analysis in the smartphone domain.

In addition, Fig. 2 shows the overall results of frequency relating to the polarity of 25,249 users in Twitter contents. In Fig. 2, the x-axis indicates the polarity of the topics in the smartphone domain, and the y-axis represents the frequency of the polarity.

Table 2. OWD for smartphone domain.

Polarity	Opinions of contents	# of
Positive	좋다, 간편하다, 부럽다, 감동이다, 빠르다, 편하다, and so on	384
Negative	나쁘다, 불편하다, 느리다, 어렵다, 까다롭다, 복잡하다, and so on	509

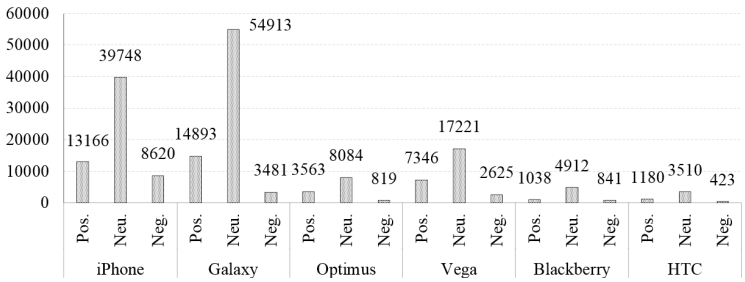


Fig. 2. Results of polarity analysis in smartphone domain.

For Intimacy analysis, in order to measure the level of intimacy among users, elements that are closely related with intimacy first need to be identified. The elements identified can then be utilized as indicators to represent the intimacy level and to reveal the differentiators between intimate and superficial relationships. For the intimacy algorithm and personalized search algorithms, the author already published in [14, 15].

4 Implementation

In this section, we describe the implementation of proposed recommender system using Twitter dataset and then show the snapshots of results. First of all, we crawled contents related topics to smartphone domain. Then we find all user who has contents in our crawled data. The Table 3 summarizes the implementation of Twitter dataset.

Table 3. Dataset of Twitter contents and users.

Domain	Topics	#of contents	#of users
Smartphone	iPhone, Galaxy, Optimus, Vega, Blackberry, HTC	259,176	25,249

For the search systems, the top ranked recommended items are the most important, because users tend to loot at only the top ranked results. In order to ranking, our PRS is focus on the intimacy algorithm based on user's relationship which is determined by ratio of sharing friends. In the implementation, we use crawled contents of Korean Twitter dataset and user relation data during one month. Whole data size of contents is about 93.1 GB, and data size of user is 12.3 GB.

Figures 3, and 4 are shown snapshots of PRS implementation. Figure 3 shows the results of keyword search such as Galaxy (Fig. 3(a)), and iPhone (Fig. 3(b)) based on intimacy. Figure 4 shows the snapshots of search frequency. Figure 4(a) indicate

search frequency of friend's with intimacy and Fig. 4(b) indicate search frequency of all user without intimacy. Our PRS implementation returns each user receive different results for the same queries.

(a) Result of keyword search(GalaxyS)

(d) Result of keyword search(iPhone)

Fig. 3. Snapshots of PRS.

(a) frequency of friends with intimacy

(d) frequency of all user without intimacy

Fig. 4. Snapshots of search frequency.

5 Conclusions

In the most existing recommender systems return a list of search results but ignore the user's specific interests, and they are impossible to personalize accurately, because they don't consider how close relationship between users is and their preference. To overcome these limitations, this paper proposed the PRS based on intimacy measurement in SNS and its implementation. Proposed approach not only can improve quality of recommendation, but also reflect individual tendency to result of recommendation. Consequentially proposed approach provides more accurate personalized results than before.

Acknowledgments. This research was supported the Next-Generation Info. Computing Dev. Program through the NRF of Korea funded by the Ministry of Science, ICT & Future Planning (2015R1C1A1A02036442). The corresponding author is Bongjae Kim.

References

1. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications, Cambridge University, NY (1994)
2. Cambria, E., Rajagopal, D., Olsher, D., Das, D.: Big Data Computing-Big social data analysis. Cambridge University, NY (2014)
3. Golbeck, J., Grimes, J.M., Rogers, A.: Twitter use by the U.S. congress. J. Am. Soc. Inf. **61** (8), 1612–1621 (2010)
4. Ronen, I., Shahar, E., Ur, S., Uziel, E., Yogev, S., Zwerdling, N., Carmel, D., Guy, I., Har'El, N., Ofek-Koifman, S.: Social networks and discovery in the enterprise. In: 32nd ACM-SIGIR, pp. 836–836 (2009)
5. Hannak, A., Sapiezynski, P., Kakhki, A.M., Krishnamurthy, B., Lazer, D., Mislove, A., Wilson, C.: Measuring personalization of web search. In: WWW, pp. 527–538 (2013)
6. Kautz, H., Selman, B., Shah, M.: Referral web: combining social networks and collaborative filtering. Commun. ACM **40**(3), 63–65 (1997)
7. Ahn, J., Brusilovsky, P., Grady, J., He, J.: Open user profiles for adaptive news systems: help or harm?. In: 16th International Conference on WWW, pp. 11–20 (2007)
8. Chen, L., Runquan, X., Xinjun, G., Lei, L., Tao, L.: Personalized news recommendation via implicit social experts. Inf. Sci. **254**, 1–18 (2014)
9. Zui, Z., Hua, L., Kun, L., Dianshuang, W., Guangquan, Z., Jie, L.: A hybrid fuzzy-based personalized recommender system for telecom products/services. Info. Sci. **235**, 117–129 (2013)
10. Yoo, D.: Hybrid query processing for personalized information retrieval on the semantic web. Knowl. Based Syst. **27**, 211–218 (2012)
11. Mehrdad, J., Norwati, M., Nasir, S., Ali, M.: WebPUM: A Web-based recommendation system to predict user future movements. Expert Syst. Appl. **37**, 6201–6212 (2010)
12. Harris Interactive Survey Result. http://www.harrisinteractive.com/NewsRoom/HarrisPolls/abid/447/mid/1508/articleId/403/ctl/ReadCustom%20Default/Default.aspx

13. Andreevskaia, A., Bergler, A.: Mining WordNet for fuzzy sentiment: sentiment tag extraction from WordNet glosses. In: EACL (2006)
14. Seol, K., Kim, J.D., Baik, D.K.: Common neighbor similarity-based approach to support intimacy measurement in social networks. J. Inf. Sci. **42**(2), 1–10 (2015)
15. Seo, Y.D., Kim, J.D., Baik, D.K.: PReAmacy: personalized recommender algorithm based on social network service. KIISE **41**(4), 209–216 (2014)

Measuring Similarity Between Graphs Based on Formal Concept Analysis

Fei Hao, Dae-Soo Sim, and Doo-Soon Park(✉)

Department of Computer Software Engineering,
Soonchunhyang University, Asan, Korea
{fhao, parkds}@sch.ac.kr

Abstract. Graph, an important information organizational structure, is commonly used for representing the social networks, web, and other internet applications. This paper tackles a fundamental problem on measuring similarity between graphs that is the essential step for graph searching, matching, pattern discovery. To efficiently measure the similarity between graphs, this paper pioneers a novel approach for measurement of similarity between graphs by using formal concept analysis that can clearly describe the relationships between nodes. A case study is provided for demonstrating the feasibility of the proposed approach.

Keywords: Graph · Similarity · Formal concept analysis · Social networks

1 Introduction

Recent advancement of large-scale graph theory techniques and ubiquitous computing paradigm enrich the mining and analysis of graph data and other complex networks system, such as Protein interaction networks [1], social networks [2] and transportation networks [3]. Therefore, understanding the internal topological structure of networks is benefit to obtain the deep insights and knowledge from the graph.

Similar subgraph matching refers to finding the subgraph structures with similar topological structure on the basis of isomorphism. In many practical applications, a given application issue is normally transformed into a graph, and then measure the similarity degree between graphs. This paper focuses on the similarity degree evaluation between graphs. There have been many existing research works on graph similarity measurement. One widely used approach, named GED is to sum the cost of elementary operations: node substitution, node insertion. However, this method is an NP-hard in general and its main shortcoming is the exponential computational complexity in terms of the number of graph edit vertices [8]. Yan et al. [9] proposed a feature-based approach for similarity search in graph structures. They used indexed features in graph database to filter graphs without performing pairwise similarity computation.

This paper is the first work on similarity measurement between graphs by using formal concept analysis. Therefore, it is a pioneering research which can bridge the gap between graph mining and soft computing. The highlights of this paper lie in presenting an efficient measurement approach for similarity between graphs. (1) First, we

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_112

construct the formal context for given graphs according to Modified Adjacency Matrix; (2) Then, the corresponding formal concept lattices are generated. (3) Obtain the similarity between graphs since the similarity between the generated formal concept lattices is equivalent to the similarity between graphs.

The rest of this paper is organized as follows. Section 2 describes the addressed problem and provides a big picture of solution. The definition of similarity between formal concept lattices is presented in Sect. 3. Section 4 elaborates the detailed approach for calculating the similarity between graphs via a case study. Conclusions is given in Sect. 5.

2 Problem Definition and Solution Idea

Before presenting our problem statement, several basic definitions such as graph, similarity are firstly provided. Then, the formalism of addressed problem is presented. Regarding to the problem, a solution

Definition 1 (Graph). In graph theory, a graph is mathematically formalized with a pair G = (V, E), where V denotes the set of vertices and E indicates the set of edges, formed by pairs of vertices. The number of vertices V and edges E are called as the cardinality of V and E, are also commonly represented by n and m or |V| and |E|.

Adjacency is defined for vertex and edge pairs: two vertices u and v are adjacent if and only if e_{uv} is an edge of the graph, while two edges e_A and e_B are adjacent if they have an endpoint in common.

Problem Statement (Graph Similarity). Given two graphs $G_1(n_1, e_1)$ and $G_2(n_2, e_2)$, with possibly different number of nodes and edges, and the mapping between the graphs' nodes, this problem is to find an algorithm to calculate the similarity of two graphs, denoted as *sim(G1, G2)* and returns a measure of similarity that captures intuition well.

Solution Idea: Different from the other existing approaches for calculating the graph similarity by using the structural features, the solution idea of this paper is to evaluate the similarity between graphs based on Formal Concept Analysis. The specific technical steps are shown as follows:

Step 1: Represent the given graphs with the formal contexts according to our previous works [4–6].
Step 2: Build the formal concept lattices for the above constructed formal contexts.
Step 3: Calculate the similarity between the formal concept lattices.
Step 4: Return the above obtained similarity to the resulting similarity between graphs.

3 Similarity Between Formal Concept Lattices

This section mainly presents the approach for evaluating the similarity between formal concept lattices. Let us revisit the methodology of formal concept analysis.

Definition 2 (Formal Context). A formal context is organized as a 3-tuple *K = (O, A, I)* with *O* and *A* indicating the objective set and attribute set respectively, and $I \subseteq O \times A$ denotes a binary relation between objective and attribute. For example, $o \in O$ and $a \in A$, $(o, a) \in I$ is interpreted as objective *o* has the attribute *a.*

Definition 3. In FCA methodology, two key operators are defined, for $X \subseteq O$, we define a set of common attributes of *X*, $X^{\uparrow} = \{a \in A | (x,a) \in I, \forall x \in X\}$; and for $Y \subseteq A$, we also define a set of common objectives of *Y*, $Y^{\downarrow} = \{o \in O | (o,y) \in I, \forall y \in Y\}$.

Definition 4 (Concept). In a formal context *K = (O, A, I)*, for $X \subseteq O$, $Y \subseteq A$, if $X^{\uparrow\downarrow} = Y$, then this pair *(X, Y)* is called as a concept where X, Y are the extent and intent of the concept.

Definition 4 (Concept Lattice). In a formal context *K = (O, A, I),* a concept lattice *L (O, A, I)* is defined that concepts organized according to a special hierarchical partial order.

After detailed presentation of the preliminary knowledge of FCA methodology, a similarity degree function between concept lattices is defined as follows,

Definition 5 (Similarity Degree Function) [7]. Let L_A, L_B be the concept lattices, the similarity degree is defined as average value of similarity degree between the nodes in L_A, L_B, thus it is formalized as follows,

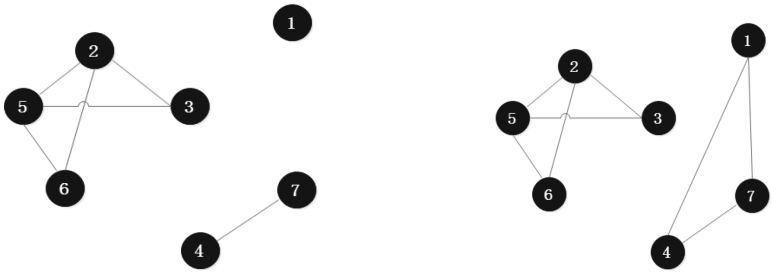
$$sim(L_A, L_B) = \frac{\sum_{C_i \in L_A} sim(C_i, L_B)}{n}$$

where $sim(C_i, L_B) = \max(\frac{\sum_{l \in R_i} sim(C_i, l)}{n})$, R_i indicates the path set which describes the concept C_i.

4 The Proposed Approach and Case Study

In this section, we technically elaborate the proposed approach for measuring the similarity degree between graphs based on formal concept analysis. As mentioned before in the solution idea, a case study is presented for illustrating the working principle of the proposed approach.

Example 1: Given two graphs g_1, g_2, the visualization of these two graphs are shown in Fig. 1. Both graphs include 7 nodes, however the topological structure is different. The target of this case is to return the similarity degree between g_1, g_2, i.e., *sim(g1, g2).*



(a) g1 (b) g2

Fig. 1. Visualization of two given graphs

4.1 Constructing the Formal Contexts for Graphs

By using the construction approach [4], the formal contexts are easily obtained as follows (Tables 1 and 2).

Table 1. Formal context of graph g1

	1	2	3	4	5	6	7
1	X						
2		X	X		X		
3		X	X		X		
4				X			X
5		X	X		X	X	
6		X			X	X	
7				X			X

Table 2. Formal context of graph g2

	1	2	3	4	5	6	7
1	X			X			X
2		X	X		X		
3		X	X		X		
4	X			X			X
5		X	X		X	X	
6		X			X	X	
7	X			X			X

4.2 Building the Formal Concept Lattice

According to the formal concept lattice generation algorithm presented in [4], the formal concept lattices of two graphs are separately shown in Fig. 2.

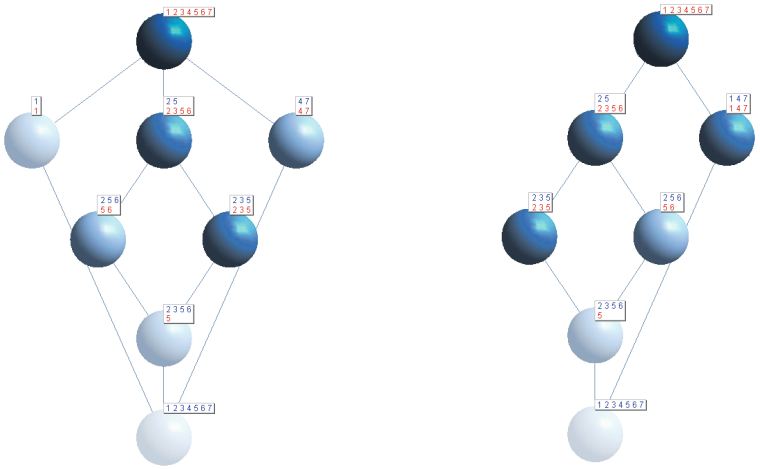


Fig. 2. The generated concept lattices for *g1* and *g2*

4.3 Calculating the Similarity Degree Between Concept Lattices

It is obviously to calculate the similarity degree between concept lattices according to Definition 5. As can be seen from Fig. 2, the difference between concept lattices of g1 and g2 lies in two concepts, i.e., there exist two special concept paths < {1,2,3,4,5,6,7},{} > -> < {1},{1} > -> < {},{1,2,3,4,5,6,7} > and < {1,2,3,4,5,6,7}, {} > -> < {1,4,7},{1,4,7} > -> < {},{1,2,3,4,5,6,7} > on *L(g1)*. Hence, the similarity between two lattices is

$$sim(L(g_1), L(g_2)) = \frac{\sum_{C_i \in L(g_1)} sim(C_i, L(g_2))}{3} = 2/3 * 1 = 0.667$$

After obtaining the similarity degree between concept lattices, we can say that it is equivalent to similarity between graphs. In another words, the similarity between g1 and g2, denoted as *sim(g1, g2)*, shown in Example 1 is 0.667.

5 Conclusions

Motivated by many promising applications and objectives, this paper aims to measure the similarity degree between two graphs, a novel formal concept analysis based evaluation approach is proposed. This approach firstly constructs the formal contexts for given two graphs, then builds the corresponding formal concept lattices of them, finally, a similarity degree function for concept lattice is defined and adopted for measuring similarity of graphs. The case study is also conducted for demonstrating the feasibility of the proposed approach. Importantly, our approach can clearly characterize the relationship between nodes and further return the similarity between graphs by calculating the similarity between nodes.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) (IITP-2015-IITP-2015-H8601-15-1009) supervised by the IITP (Institute for Information & communications Technology Promotion) and Basic Science Research program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2014R1A1A4A01007190).

References

1. Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Li, G.: Topological structure analysis of the protein–protein interaction network in budding yeast. Nucleic Acids Res. **31**(9), 2443–2450 (2003)
2. Ghosh, S., Ganguly, N.: Structure and evolution of online social networks. In: Panda, M., Dehuri, S., Wang, G.-N. (eds.). ISRL, vol. 65, pp. 23–44Springer, Heidelberg (2014). doi:10.1007/978-3-319-05164-2_2
3. Tong, L., Zhou, X., Miller, H.J.: Transportation network design for maximizing space–time accessibility. Transp. Res. Part B Methodol. **81**, 555–576 (2015)
4. Hao, F., Min, G., Pei, Z., et al.: K-clique communities detection in social networks based on formal concept analysis (2015)
5. Hao, F., Park, D.S., Min, G., Jeong, Y.S., Park, J.H.: k-cliques mining in dynamic social networks based on triadic formal concept analysis. Neurocomputing (2016). doi:10.1016/j.neucom.2015.10.141
6. Hao, F., Yau, S.S., Min, G., et al.: Detecting k-balanced trusted cliques in signed social networks. IEEE Internet Comput. **18**(2), 24–31 (2014)
7. Hao, F., Zhong, S.: Tag recommendation based on user interest lattice matching. In: 2010 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), vol. 1, pp. 276–280. IEEE (2010)
8. Zeng, Z., Tung, A.K.H., Wang, J., Feng, J., Zhou, L.: Comparing stars: on approximating graph edit distance. Proc. PVLDB **2**(1), 25–36 (2009)
9. Yan, X., Zhu, F., Yu, P.S., Han, J.: Feature-based similarity search in graph structures. Proc. ACM Trans. Database Syst. **31**, 1418–1453 (2006)

TEXAS2: A System for Extracting Domain Topic Using Link Analysis and Searching for Relevant Features

SangWon Hwang(✉), YongSeok Lee, and YoungKwang Nam

Department of Computer Science, Yonsei University, Wonju, South Korea
{arsenal, yknam}@yonsei.ac.kr, asapfast@naver.com

Abstract. It is very important to understand the domain topic of software to maintain and reuse it. However, the continual development and change in its size makes it difficult to understand it. To solve this problem, researches have been recently conducted to extract the domain topic using various information search techniques such as LDA, with the researches on LDA-based techniques being especially active. However, since only unstructured information such as an identifier or note is used in most research, without including structured ones like information calling, problems in which extracted topics are different from the characteristics of the program can occur. In this paper, we propose a method to generate documents and extract topics using both structured and unstructured information. We also generate indexes based on the frequency of the identifier of the source code, and propose a system that extracts an association rule based on the simultaneous generation of the method. We as well establish a system that provides highly reliable search results to user queries by combining domain topics, indexes with scores, and the association rule information. Consequently a TEXAS2 system for this study was established and confirmed a high user satisfaction on search results to the queries in a performance test.

Keywords: Topic modeling · LDA · Latent Dirichlet Allocation · Feature location · Feature identification · Software reuse · Sequential mining

1 Introduction

As the size of projects become larger and their application systems become diverse, structures of software get more complicated. Moreover, once the software is developed, it is necessary to keep the software updated through continuous modifications, improvements, debugging, and performance enhancements and these tasks require a good understanding of the software system. The most important step of a good understanding of the software system is to determine the location of the implemented features within the source code, or "Feature" or "Concept Location" [1]. If an engineer is not familiar with the software system, he/she will be required to read through the entire code to find the location of the program topic (Feature/Concept). This method for reviewing the entire structure of a large-scale system that has hundreds of classes and methods is however inappropriate and time-consuming. In general, there are many methods for identifying domain topics for mid or larger projects, such as call graph,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_113

control flow and data flow, which are program analysis techniques [2]. Nonetheless, although these methods help engineers understand the structural information, such methods are less likely to help them understand the domain topics. On the other hand, many efforts have of recent seen LSI or LDA algorithms applied to non-structured data in a source code to find the topics [3–6]. These methods can locate domain topics by finding multiple word-sets about the project (one word-set is equivalent to one topic) and estimating the possibility of the project to be included in those word-sets (topic distributions). The engineer can then name each topic to estimate or extract domain topics. The estimation might not be very accurate however because the topic distribution in these methods is very small, and the correlation between the topics is not high enough.

In our study, TEXAS2 (Topic EXtract And Search System) was designed and implemented to extract, save, and search a topic of a source code effectively. TEXAS2 first generates text documents corresponding to each source code of software, and thus the software will consist of document sets. LDA topic modeling is then applied to the document sets. Each source code contains a class that has a higher possibility of being a topic than other words, and this is used when a text document is generated for the analytical purpose after the importance of the identifier in the class is calculated using PageRank. After the API (method, function) call sequences are extracted using sequential pattern mining algorithm [7], the weight of each identifier is set according to its API call information. Furthermore, indexes are created by Lucene [8] for all the generated information and source codes, and a search engine is developed using Solr [9] so that users can search for the information that they need.

2 Related Work

Blei, A. Ng and M. Jordan [10, 11] proposed LDA (Latent Dirichlet Allocation) algorithm, a probability model, that estimates topics in text documents in natural language. In their method, the algorithm can estimate the topic discussed in a specific document by analyzing the word count distributions in the document and comparing them with the distributions of the word counts that were already studied and known for each topic.

The document in LDA contains several topics, and only word count is an important variable, but not the sequence of the words. The main purpose of LDA is to estimate the topics that are mainly discussed in a document by comparing the word count

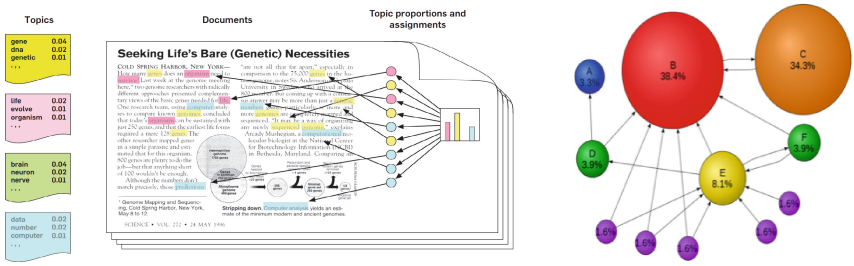


Fig. 1. The concept of LDA(left) and the concept of PageRank(right)

distribution in the document with the predefined word count distribution found for each topic. In Fig. 1, the left portion of the diagram illustrates the LDA.

Sergey Brin and Lawrence Page [12] proposed PageRank and this algorithm ranks the documents by weighing them according to the importance of the documents, and these document sets have a WWW link structure. The right portion in Fig. 1 shows the result when the PageRank algorithm is applied to a web page graph. Each node represents a web page, and an edge indicates that one web page includes a link to another web page. The PageRank algorithm uses a method that normalizes the number of links to each node, and uses the following equation:

$$PR(A) = (1-d)/N \; + \; d(PR(T_1)/C(T_1) + \ldots + PR(T_n)/C(T_n)). \tag{1}$$

where PR is PageRank score, PR(A) is a score of a web page "A", d is a damping factor, N is the number of pages, Tn is the page that points to the web page "A", and C (Tn) is the number of links that Tn has. The damping factor ranges from 0 to 1, and it indicates the probability that a user moves to another web page through a link on the current web page. This number is normally determined through an experiment, and 0.85 is widely used. Overall, the PageRank score of a specific web page is calculated by adding up the normalized PageRank scores of all web pages that point to the specific web page, and a PageRank score is then allocated to each web page.

Lucene is a high-performance and expandable search library written in Java and provided by Apache Software Foundation. With this software, documents can be indexed and searched, and also full-featured text search is possible. This software, however, only provides the API for the indexes and searches, but does not directly collect documents nor answer any user queries. Therefore, users collect and index the documents by their own or use a crawler, and search engines use Lucene-based search platforms including Solr.

Solr is an open source enterprise search platform built on Apache Lucene. This platform supports real-time indexing and documents of various formats and provides very strong features including a full-text search.

Apriori algorithm is one of the earliest developed influential algorithms, and very widely used. This algorithm finds a strong association between one item-set and another item-set. For example, when one transaction is defined as an item-set that customer purchases and DB as a set of transactions for a certain period, we can express a rule that "if a customer purchases a diaper, the person also purchases a beer" as "diaper=>beer [10 % support]". This 10 % support means that in a given DB, 10 % of customers purchase a diaper and a beer at the same time, and a confidence level of 80 % means that 80 % of customers who purchase a diaper also purchase a beer. In the association rule search, the user can input the appropriate support and confidence levels to discover a correlation between products among transactions that have occurred [13].

3 TEXAS2 System

Figure 2 shows the entire system architecture of the TEXAS2. This system consists of a preprocess of source codes, call graph generator, topic modeling, source code indexer, association rule generator, and search engine part.

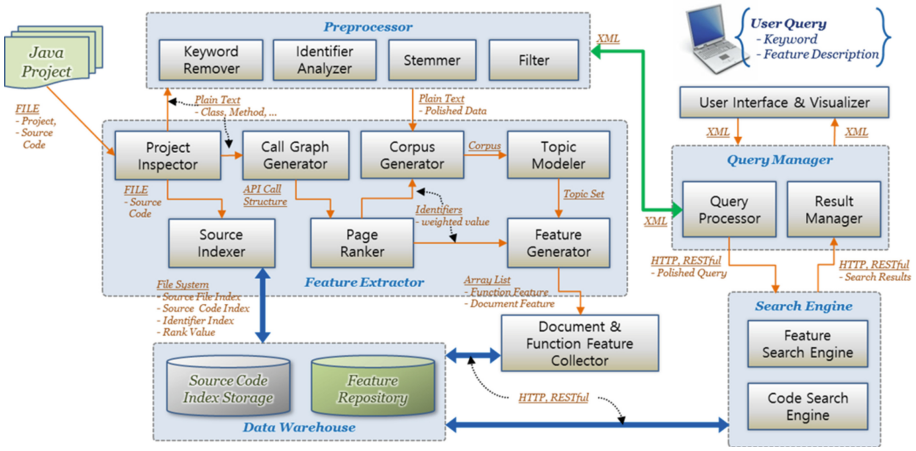


Fig. 2. The TEXAS2 system overview

The source code is different from a conventional document because it is a structural document containing keywords, reserved words, compound words and stop words, so this difference makes the source code difficult to use for the topic modeling. In the pre-process stage, the all words(class, method, comments, …) that cannot be used for topic modeling are eliminated, such as the programming keywords(ex: private, public), the stop words(ex: is, the), camel-cased(ex: drawCircle) and underscored(draw_circle) compound words are properly treated, and original words are extracted by Porter's Stemming algorithm. Furthermore, widely-used words including "get" and "set" are filtered to generate a corpus that can be used in the topic modeling.

The call graph generator calculates the counting of the calls between codes in the program, and the extractions and calls of the called classes. In this case, the classes are defined as nodes and the relations between method calls within the classes are defined as edges to generate a graph. This graph is then used to score each node using PageRank algorithm. Based on the score of each node, an identifier of the node is added to a corpus that is generated from the pre-process stage.

In the topic modeling process stage, LDA-based topic modeling which is applied by the generated corpus, and the distribution and word sets of random topics are found for each document in the document sets. In addition, users confirm the outputs and name a topic for each word set. If no special work is to be performed, the extracted word is used as the topic name.

In the source code indexing stage, each code is indexed. Based on the words within the code, it is weighed using a weighing algorithm such as TF/IDF, scored, and the score index is stored.

We extract an association rule of the method in the association rule generation stage. For example, we generate a combination rule (setCurrentDirectory(), showOpenDialog (), getSelectedFile()) in which 'setCurrentDirectory()', 'showOpenDialog()' and 'getS-electedFile()' are called sequentially, and store both its support value and confidence

level. The method with a higher frequency of the simultaneous appearance has a higher possibility of use by users sequentially, and this affects the ranking of the search results.

In the search stage, the search on the user's query is performed using the topics, indexed source code and combination rule information. For this case, the ranking is determined by the topics of the project, index score generated from the indexing engine, and the confidence level assigned to the association rules.

4 Result

The TEXAS2 GUI was developed by C#, and MALLET toolkit was used for topic modeling. Indexer was established using LUCENE 6.1 and search engine was constructed with Solr 6.1. To generate the association rule, we developed a JAVA parser based on JAVA 1.8 version by JavaCC. Topic modeling, search engine and core module for integrated interface with UI were developed by JAVA. For the initial variables used for the test, the number of topic sets was 20, and the number of words for each topic was 20 as well. The repetition of the sampling was conducted 500 times, and the confidence of the association rules was 85 %. Also 25 Open Source Systems including JEdit, JHotDraw were used for the document of Topic Modeling. 10 people participated in the experiment and the results were compared with MALLET. The percentage of words extracted from each project that can be used for the domain topics, TEXAS2 was 74 %, while MALLET was 35 %, indicating that the system used in this study is efficient. Also to measure the effectiveness of the system, we tested on (1) LUCENE search system standalone, (2) LUCENE search system + the search system based on topic modeling, (3) LUCENE search system + the search system based on topic modeling + the search system based on association rule information(our system). The satisfaction rates of search result were (1) 73 %, (2) 80 %, (3) 86 % respectively. Based on the results, (3) was the most effective technique in this experiment.

5 Conclusions

In this study, we proposed a method for topic modeling by extracting weighted value from function call information to compensate the deficiencies occurring when extracting software domain topics based on unstructured text. We used simultaneously generated information of the method to extract reliable association rule information, and established a search system that reflects it. As a result, users were highly satisfied when they searched for a project or code information using this information. It is expected that the system will become more complete as its reliability is verified with more projects in the future.

Acknowledgement. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Science, ICT & Future Planning(NRF-2014M3C4A7030505).

References

1. Antoniol, G., Guéhéneuc, Y. G.: Feature identification: an epidemiological metaphor. IEEE Trans. Softw. Eng. **32**(9), 627–641. IEEE Press, New York (2006)
2. Karrer, T., Krämer, J.P., Diel, J., Hartmann, B.: Stacksplorer: call graph navigation helps increasing code maintenance efficiency. In: Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, pp. 217–224. ACM, New York (2011)
3. Maskeri, G., Sarkar, S., Heafield, K.: Mining business topics in source code using Latent Dirichlet Allocation. In: Proceedings of the 1st India Software Engineering Conference, pp. 113–120. ACM, New York (2008)
4. Alenezi, M.: Extracting high-level concepts from open-source systems. Intl. J. Softw. Eng. Appl. **9**(1), 183–190 (2015). SERSC, Tasmania
5. McBurney, P.W., Liu, C., McMillan, C., Weninger, T.: Improving topic model source code summarization. In: Proceedings of the 22nd International Conference on Program Comprehension, pp. 291–294. ACM, New York (2014)
6. Savage, T., Dit, B., Gethers, M., Poshyvank, D.: Topic XP: exploring topics in source code using Latent Dirichlet Allocation. In: IEEE International Conference on Software Maintenance, pp. 1–6. IEEE Press, New York (2010)
7. Slimani, T., Lazzez, A.: Sequential mining: patterns and algorithms analysis. Intl. J. Comput. Electron. Res. **2**, 639–647 (2013)
8. Apache Lucene. https://lucene.apache.org/core/
9. Apache Solr. https://lucene.apache.org/solr/
10. Blei, D., Ng, A., Jordan, M.: Latent Dirichlet Allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003). MIT Press, Cambridge
11. Blei, D.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012). ACM, New York
12. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. J. Comput. Netw. ISDN Syst. **30**, 107–117 (1998). Amsterdam
13. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: Proceedings of the 20th VLDB Conference

Ubiquitous Computing for Cloud Infrastructure to Mobile Application in IoT Environment

DongBum Seo[1], Keun-Ho Lee[2(✉)], and You-Boo Jeon[3]

[1] Department of Visual Information Processing,
Korea University, Seoul 02841, Korea
treeline@korea.ac.kr
[2] Division of Information and Communication, Baekseok University,
Cheonan-si, Chungcheongnam-do 31065, Korea
rootl004@bu.ac.kr
[3] Industry-Academic Cooperation Foundation, Soonchunhyang University,
Asan-si, Chungcheongnam-do 31538, Korea
jeonyb@sch.ac.kr

Abstract. The growth of the Internet of Thing (IoT) ability up all types of service driven to ubiquitous cloud infrastructure different access methods are analyses to understand new message protocols that are used ubiquitous IoT mobile application environment that presented the cloud computing platform for mobile application. It supports a combined architecture of ubiquitous and cloud computing which provides a how device can grow in intelligence, interoperability with other IoT environment, system and service. The Cloud Infrastructure for ubiquitous computing environment mobile application (CI-UCEMA), which consist of three layers it Cloud Service Layer (CSL), M2M Service Layer (MSL) and Ubiquitous Service Layer (USL). The M2M consists of IoT Services layer (MSL) will involve a decrease in complexity of both the improvement and controlling of IoT systems. Realizing the full potential of the Internet of Thing requires that we change how we view and build ubiquitous environment which provide the core foundation of service.

Keywords: Ubiquitous computing · Cloud computing · Cloud infrastructure · M2M mobile application

1 Introduction

The Internet of Thing (IoT) follows the Ubiquitous environment and will have an equally wide ranging impact, with the potential to "things". It will change everything, including how things are made and used and even our environment. Ubiquitous computing will feature intelligent device at the edges of networks. Further, Ubiquitous networks will not be connected to just the Internet most of solutions, will also be connected to other ubiquitous environments, as well as the Internet. This will result in a composite, decentralized environment with intelligence at the edge and on the Internet, and in the cloud computing environment to ubiquitous computing. As the number of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_114

connected things increases to ubiquitous computing that service are defined and built delivered will change increased. We defined into the car blackbox connected from time of purchase and always connected to the infrastructure and user through such ubiquitous new service will be enabled. These are now appearing in all areas and industry sectors, including automation vehicle telemetric and car information transfer to another place PC or Smartphone. New composite solutions created by connecting ubiquitous real value, which can provide both reduce costs and new revenue opportunities to business across all ubiquitous environments to IoT.

2 System Architecture

In this approach, we describe an architecture of our system so called CI-UCEMA (Cloud Infrastructure for Ubiquitous Computing Environment Mobile Application), which consists of three layers: Cloud Service Layer (CSL), M2M Service Layer (MSL) and Ubiquitous Service Layer (USL).

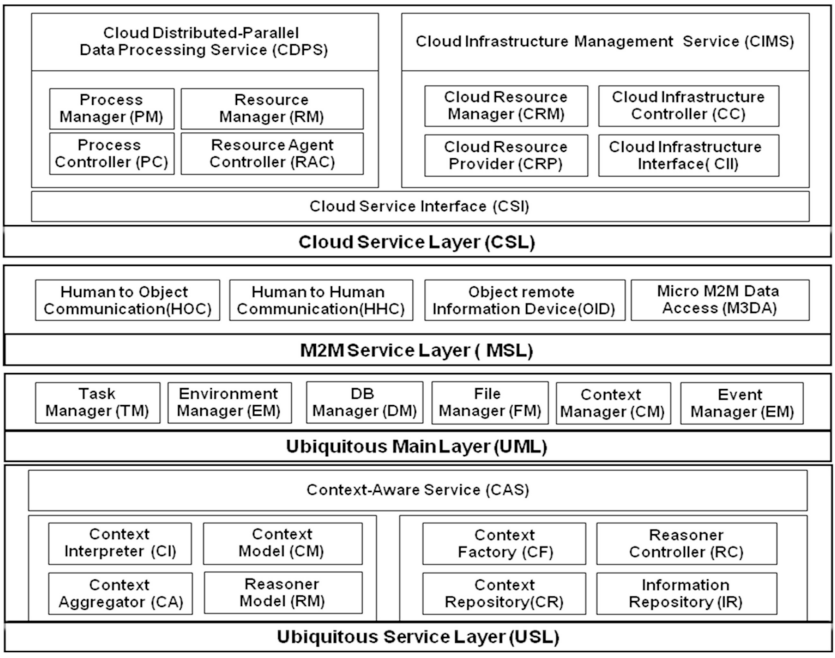


Fig. 1. System architecture of CI-UCEMA

Our analysis of these architecture as shown in Fig. 1. The CI-UCEMA infrastructure supports all functions that it can provide internally with each layer operating.

2.1 Cloud Service

The most developed prior work in this area is the Cloud Service Layer consists of two components: Cloud Distributed Parallel data processing Service (CDPS) and Cloud Infrastructure Management Service (CIMS). CDPS provides an agent based processing platform for executing various distributed models through four components: Process Manager (PM), Process Controller (PC), Resource Manager (RM) and Resource agent Controller (RC) [1]. This layer provides a software execution IoT environment, and manages execution of user mobile application.

2.1.1 Cloud Distributed Parallel Data Processing Service

CDPS supports various distributed or parallel models based on the connection pattern between PC and RC created by PM and RM respectively. A resource agent is created in each VM, and connected to RC in CDPS, performing data transfer, program execution and running for running mobile applications through RC in CDPS [2–4].

2.1.2 Cloud Infrastructure Management System

Also, CIMS offers cloud infrastructure controller which supports a unified interface for various cloud infrastructure via cloud infrastructure interface so that mobile applications may be launched on any cloud environment. PM runs a set of processes by creating their corresponding PC each of which in turn manages a resource agent in VM through RC. Resource Agent transfers a user data for applications and configurations from Data Repository to the VM resource, configures an initial status before running mobile application. After all allocated VM resources are configured, each Resource Agent runs mobile applications by order of Process Controller, and monitors the utilization of VM resources and running status of mobile application. CIMS provides on-demand VM allocation for dynamic execution of mobile processes through Cloud Resource Manager which receives resource requirements for mobile application from USL, and launches VM instances via Resource Provider.

2.2 M2M/IoT Service

M2M Service will involve a decrease in complexity of both the improvement and controlling of Ubiquitous M2M/IoT systems. This can be distinguished by improved interoperability and, carefully related, the improvement and use of normal. Security will also remain a grassroots regard as device management and provisioning are accomplished [5]. M2M Service layer (MSL) is responsible for shielding the user from the underlying complexity and variability through self-tuning environment by mobility and adaptation in CDPS and CIMS. MSL Components has Human to Object Communication (HOC), Human to Human Communication (HHC), Object remote Information Device (OID) and Micro M2M Data Access (M3DA). HOC has something concerned communications with PC, TV, PDA, Wearable PC, Mobile Phone and LTE Black Box. HHC supports services concerned with making same Human to Human Communications environment in everywhere [6, 7].

2.2.1 Human to Object Communication

Ubiquitous computing environments enable us to process humongous amounts of information that, for example enable us to detect the locations of indoor objects with position sensors such as obtains vast amounts of environmental sensor data on M2M/IoT environment. We use relationships between environmental objects to refer to an object in human to human interactions. It is useful for computers to treat relationships between objects in human to Smartphone or computer interactions.

2.2.2 Human to Human Communication

The Human to Human Communications initially tend to treat objects as human and think they have a object. It is often suggested that the transportation and telecommunications are strongly coupled, or complementary. The telephone is an important means of making appointments and therefore enabling and justifying more HHC, but at the same time HHC increases the need for communications.

2.2.3 Object Remote Information Device

Object remote Information Device (OID) on introduction to a new, emerging communication standard with high relevance for the growing Machine to Machine (M2M) industry and the Internet of Things (IoT) as a whole. As an open interface remote management protocol standard between M2M appliances and server-side M2M platforms optionally Cloud based, it leads to a decoupling of both sides, thereby enabling greater independent innovation of M2M devices and M2M platforms.

2.2.4 Micro M2M Data Access

M3DA is an open-source [6]; royalty-free protocol Specification to be released in open source. We generate documentation with more high level specification for Mihini agent features and Framework modules. We use a Koneki Execution Environment to help developers using our APIs with Koneki tools. This Execution Environment will enable code completion, content assist, and documentation pop-up while using Mihini APIs. The Enterprise application must use robust certification to authenticate that it can principal data from the CIMS. Also M2M CIMS is that data coming from both used device and enterprise application can be principal platform.

2.3 Ubiquitous Service

The Ubiquitous Service Layer [8, 9] consists of two components: Context-Aware Service (CAS) and Ubiquitous Man Service (UMS). It is responsible for shielding the user from the underlying complexity and variability through intelligent context-aware infrastructure and automatic ubiquitous service in the self-tuning environment for mobility and adaptation for ubiquitous application. In mobile computing, whenever the user moves from one place to another, the tasks such as cloud resource allocation, environment variable configuration, and remote service execution can be automatically set up and performed by UMS and CAS [10].

2.3.1 Context-Aware Service

CAS provides a context-aware infrastructure which supports the gathering of context information from different sensors and the delivery of appropriate context information to mobile applications. Context Interpreter gathers contextual information from sensor network, and Context Reasoner provides high level contexts for mobile application through reasoning based on ontology-based Context Model [11], storing the resulting contexts in Context Repository. UMS provides on-demand automatic service execution environment for mobile application by interacting with CSL through Cloud Service Interface (CSI).

2.3.2 Ubiquitous Main Service

The Ubiquitous Main Layer [3, 11] (called uMain) is responsible for shielding the user from the underlying complexity and variability through self-tuning environment by mobility and adaptation which are weak points in CDPS and CIMS. Whenever the user moves from one place to another, the tasks and devices such as cloud authentication, environment variables, video/audio device or large display are automatically set up or executed, keeping their environment.

3 Demand Model

For a particular ubiquitous service provider, the demand d_n for its services decreases as its response time rt_n increases; on the other hand, it increases with the increase of its competitor response time rt_m, for $m \cdot n$. The analogous relationship holds for loss probabilities, but then, d_n increases with decrease of lr_n and increase of lr_m, for $m \cdot n$. The reverse relationship holds for mobility probabilities, but then, d_n increases with increase of mt_n and decrease of mt_m, for $m \cdot n$. We now consider the case where the demand function (d_n) is linear in all QoS parameters [7]. That is,

$$d_n(rt, lr, mt) = \left\{ \sum_{m \subseteq G, m \neq n} \alpha_{nm} rt_m - \theta_n rt_n + \sum_{m \subseteq G, m \neq n} \beta_{nm} lr_m - \delta_n lr_n - \sum_{m \subseteq G, m \neq n} \chi_{nm} mt_m + \eta_n mt_n \right\} \tag{1}$$

Where $\alpha_{nm}, \theta_n, \beta_{nm}, \delta_n, \chi_{nm}, \eta_n$ are constants, and mt_n is mobility value at nth cloud service provider. Here we make some minimal assumptions regarding the demand function.

4 Experiment

We intend to evaluate application performance through benchmarking for a number of applications and simulating various environment provisioning strategies to find optimal strategy that shows high performance and service quality comparative evaluation of performance with existing cloud PaaS [1, 12] can be a good choice of evaluation. Our CI-UCEMA prototype is hosted on a server with Resource Agent is pre-installed in the

Table 1. Runtimes compare on CPUs EC2 and CI-UCEMA infrastructure

Division	Subdivision	One Month	Next Month
Status of Equipment (SOE)		59	59
Change Management System (CMS)		14	16
Fault Management (FM)		6	5
Capacity Management (CM)	Internet Data Center (IDC)	33.30 %	17 %
	SystemLoad (SL)	0.3	0.2
	Storage	20 %	20 %
Backup Management (BM)		11	9
Service Availability Management (SAM)	TYPE 1	99.90 %	99.25 %
	TYPE 2	99.90 %	99.25 %
Remote Service (RS)	Contents Delivery Service (CDN)	10 %	14 %
	Security Control (SC)	0	0
	Security Vulnerability Analysis (SVA)	97.80 %	94.50 %
	Computer Viruses (CV)	0	0
Propose Improvement (PI)		2	2
Task Level Assessment (TLA)		73	75

customized VM bundle image and started automatically after booting of VM instance (Table 1).

In the LTE CAR Black Box video 3G and LTE comparison as shown in Fig. 2, the number of subscribers: 500 standard terminal Two types of services conducted by cost type review type1: The vehicle position control type2: vehicle location control and LTE car accident video services cost review preconditions excluding local SMS or mail

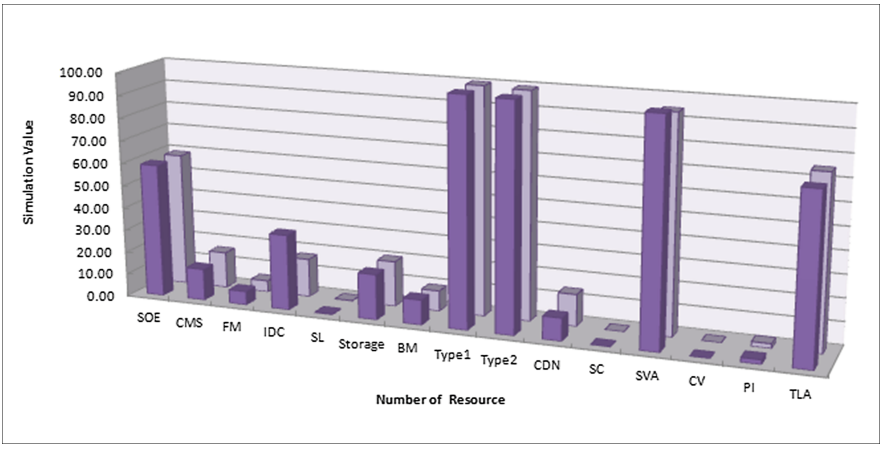


Fig. 2. Simulation value of number of resource

solutions and GIS license fees cost Smartphone applications. The development is including correspondence provided by terminal applications. Porting costs are excluded except for service operations and after service costs. We ran the OpenMP version of NAS Parallel Benchmarks NPB (NPB3.3-OMP) Class on a high CPU extra large instance and on a compute node on the CI-UCEMA cluster. Figure 2 shows the simulation value of each of the programs in the benchmark. In general we get a performance degradation of approximately 8 % to 18 % for the programs running on the EC2 nodes compared to running them on the CI-UCEMA cluster compute node. This value is shown in the overlaid line chart in Fig. 2.

5 Conclusion

We have presented the ubiquitous computing environment on cloud computing platform for mobile application. It supports a combined architecture of ubiquitous and cloud computing which provides a powerful framework for ubiquitous mobile applications which requires high performance. Also, our system infrastructure offers cloud controller which supports a unified interface for various ubiquitous infrastructure via cloud infrastructure interface so that mobile applications may be launched on any ubiquitous environment. In this new approach for local cloud infrastructure different access methods are analyses to determine their security aspects. It is important to understand new message protocols that are used for ubiquitous M2M and IoT mobile application environment. More suggested extensions to the model are discussed and general directions for future research are given for cloud infrastructure to ubiquitous M2M and IoT environment mobile application. In the comparison result CI-UCEMA better than the normal system without cloud service as the scale of simulation is increasing, although the time consumption of initialization has an effect on the state of small resource. CI-UCEMA can utilize abundant computing power and adapt for various environments, as well as provide convenient user interface to mobile application in IoT environment.

Acknowledgments. This research was supported by Korea Sanhak Foundation in 2016, the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. IoT Services & Frameworks: http://iot.eclipse.org/frameworks.html
2. Sundmaeker, H., Guillemin, P., Friess, P., Woelfflé, S.: Vision and challenges for realizing the internet of things. In: CERP-IoT - Cluster of European Research Projects on the Internet of Things, Book, pp. 43–189 (2010)
3. Seo, D., Jung, I.-Y., Park, J.H., Jeong, C.-S.: Ubiquitous mobile computing on cloud infrastructure. In: Jeong, H.Y., Obaidat, M.S., Yen, N.Y., Park, J.J.(Jong Hyuk) (eds.). LNEE, vol. 279, pp. 571–576Springer, Heidelberg (2014). doi:10.1007/978-3-642-41674-3_81
4. Amazon Elastic Mapreduce. http://aws.amazon.com/elasticmapreduce

5. Griswold, W.G., Boyer, R., Brown, S.W., Truong, T.M.: A component architecture for an extensible, highly integrated context-aware computing infrastructure. In: 25th International Conference on Software Engineering, Proceedings, 3–10 May, pp. 363–372 (2003)
6. M3DA. http://www.eclipse.org/mihini/
7. Roy, N., Das, S.K., Basu, K., Kumar, M.: Enhancing availability of grid computational services to ubiquitous computing applications. In: 9th IEEE International Parallel and Distributed Processing Symposium, Proceedings, 04–08, p. 92a (2005)
8. Borthakur, D., Gray, J., Sarma, J.S., Muthukkaruppan, K., Spiegelberg, N., Kuang, H., Ranganathan, K., Molkov, D., Menon, A., Rash, S., Schmidt, R., Aiyer, A.: Apache hadoop goes realtime at Facebook. In: Proceedings of the International Conference on Management of Data (SIGMOD 2011), pp. 1071–1080 (2011)
9. Davies, N., Friday, A., Storz, O.: Exploring the grid's potential for ubiquitous computing. IEEE Pervasive Comput. **3**(2), 74–75 (2004)
10. Roy, N.: Providing better QoS assurance to next generation ubiquitous grid users. MS Thesis, University of Teaxs at Arlington, USA (2004)
11. Seo, D.B., Lee, T.D., Jeong, C.S.: Ubiquitous mobile computing on cloud infrastructure. In: Trends in Communication Technologies and Engineering Science. Lecture Notes in Electrical Engineering. Springer, Berlin (2009)
12. Sundmaeker, H., Guillemin, P., Friess, P., Woelfflé, S.: Vision and challenges for realizing the internet of things. In: CERP-IoT - Cluster of European Research Projects on the Internet of Things, Book, pp. 43–189 (2010)

What Are Learning Satisfaction Factors in Flipped Learning?

Kyung Yeul Kim and Yong Kim(✉)

Department of e-Learning, Graduate School,
Korea National Open University, Seoul, South Korea
aooskrap6920@gmail.com, dragonknou@knou.ac.kr

Abstract. Flipped learning is an effective teaching-learning method implemented with learner centered interactive learning instead of teacher centered cramming methods of lecturing. Despite that it is an effective learning method, studies of flipped learning are not sufficient. The present study analyzed factors that affect flipped learning satisfaction. According to the results of analysis, satisfaction with interactions was directly affecting learning satisfaction and indirectly affecting learning satisfaction mediated by satisfaction with assistance services. Based on the present study, flipped learning is believed to creatively help the development of learning as an effective learning method and expected to become a new alternative education method with student centered learning.

Keywords: Flipped learning · e-learning · Blended learning · Teaching method

1 Introduction

As information communication technology has been applied to the field of education utilizing e-Learning contents, along with qualitative and quantitative improvement of education, diverse changes occurred in traditional teaching-learning interactions. Flipped learning that grafted online learning methods on offline learning methods is first learning in out of classroom classes using video lectures, etc. and performing problem solving activities later in classes in classroom. Flipped learning consists of problem-based learning activities found in active activities and constructionism and direct instruction methods found from behaviorist principles for educational lectures. It uniquely mixed incompatible learning method [1]. Studies that utilized flipped learning reported that flipped learning showed effects not only for the enhancement of learner's learning efficiency but also in terms of learning satisfaction and academic achievement [2, 3].

Despite that flipped learning is attracting great attention and has value for educational utilization, negative aspects of the new learning method also exist. Teachers feel burdens in the preparation of video files to be used in learning and the classroom environment changing from lecture centered one to learning guide centered one [4]. Due to the great change of the switching from teacher centered learning paradigms which are traditional lecturing methods to learner centered learning paradigms, teachers may misunderstand that their role to teach, which is the leadership in education will disappear [5]. As for learners, in prior learning, learners cannot maintain high levels of concentration for more than one hour while they are moving [6]. Furthermore, they feel

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_115

burdens of preparation when they do self-directed learning and those who are familiar to existing traditional learning method feel the new flipped learning method inconvenient [5]. In addition, a paper was published indicating that there was no big difference in learners' performance in flipped learning and traditional learning experiments [7].

In the present study, the measurement variables of each of four types of satisfaction appearing in flipped learning were explored and direct causal relationships among latent variables that constitute satisfaction and indirect causal relationships mediated by parameters were analyzed. Therefore, attempts were made to find what the variables that must be considered for effective flipped learning and factors that affect the learning satisfaction recognized by learners.

2 Related Works

Many researchers regard that flipped learning began with the 'peer instruction' of Professor Eric Mazur in 1990 [4]. A name 'Classroom Flip' was used in a paper of J. Wesley Baker in 2000 and in 2007, two chemistry teachers Jornathan Bergman and Aron Sam, who are called pioneers of flipped learning, used YouTube [3, 8].

In the case of flipped learning, in out of classroom classes, learners are provided with opportunities to obtain learning exposure before classes in classroom. This means that students are provided with incentives for preparation for classes in classroom. Thereafter, in classes in classroom, in flexible environments somewhat not quiet, learners' active learning is maximized. Instead of lecture centered learning, the teacher plays the role of a learning guide to induce learners to study actively [4].

Flipped learning provides learning connection between out of classroom classes and classes in classroom, learners are provided with sufficient time for performance of tasks, and teachers help learners with fair evaluation and immediate feedback. While having control over their learning, learners can obtain personal help from their colleagues or teacher. Through the entire course, the newest education skills, etc. are taught to both learners and teachers, who become to naturally learn [3, 9, 10].

Satisfaction in teaching-learning is the degree to which learners' comprehensive satisfaction with learning including satisfaction with teachers, learners, learning contents, support with learning systems, and utilization of learning systems is formed and felt through actual class experience [10–13]. The definitions of variables that must be considered for learners' learning effects and satisfaction when classes are implemented with flipped learning are as shown in <Table 1> below.

The system convenience satisfaction (FSCS) means the satisfaction of nomadicity of learners when they perform pre-watched learning outside the classroom [14]. The interaction satisfaction (FIS) is the degree of satisfaction felt for interactive learning when learner centered learning is implemented. The assistance service satisfaction (FASS) is the degree of satisfaction felt by learners for the learning assistance service in flipped learning assistance systems because of the characteristics provided by the environment necessary for learning operation. The final learning satisfaction is the degree to which the desire for learning felt by the learner is satisfied and the learning satisfaction that corresponds to overall satisfaction acts as an important element of evaluation of learning performance [15].

Table 1. Operational definition of satisfaction variables in flipped learning

Variable	Operational definition
System Convenience Satisfaction (FSCS)	The degree to which learners feel convenience in learning when using the system in out of classroom class
Interaction Satisfaction (FIS)	The degree to which learning is felt beneficial by learners in classes in classroom because of the learning related knowledge and information exchange occurring between teacher-learner, learner-learner
Assistance Service Satisfaction (FASS)	The degree to which the learning assistance information such as learning information and notices provided to learners in out of classroom classes-classes in classroom is useful
Learning Satisfaction (FLS)	The degree of learning goal clarity, interesting learning, learning(content) suitability, learning(level) suitability, possibility of self-directed leaning felt by learners in out of classroom classes-classes in classroom

3 Results of Experiments and Analysis

For experiments, learning materials were provided to learners to implement out of classroom classes and constructivism based classes were implemented in classes in classroom. Data from a total of 123 learners that participated in the experiment. The data were collected using questionnaires consisting of 5-point questions made based on previous studies. The results were analyzed using SPSS and AMOS 18.0. The analyzed reliability, the means of individual measurement variables, and concept validity are as shown in <Table 2> below.

System convenience satisfaction measurement variables' total mean was identified as 3.90 and the reliability was identified as .745. When learners watch video files in advance for learning, content accessibility should be emphasized first for learners and when contents searches were not difficult and downloads were easy, learning satisfaction was shown to have been enhanced in flipped learning.

The interaction satisfaction measurement variables' total mean was identified as 4.11 and the reliability was identified as .718. Learning satisfaction in flipped learning is judged to have been enhanced when questions and answers with teachers or learners were exchanged fast and sincerely while learners were utilizing the learning of the knowledge watched in advance.

The assistance service satisfaction measurement variables' total mean was identified as 3.89 and the reliability was identified as .723. Learning satisfaction in flipped learning is judged to have been enhanced when diverse contents were provided as well as prior notices and guides while learners were conducting learning activities.

The learning satisfaction measurement variables' total mean was identified as 4.11 and the reliability was identified as .683. Learning satisfaction in flipped learning is judged to have been enhanced when the level of lectures was not too high for learning and was satisfactory to learners. All correlation coefficients obtained through confirmatory variable analyses were between .5 – .6 indicating correlations.

Table 2. Means of measurement variables, reliability, and concept validity

Variable	Division	Measurement variable	Mean/SD	Total mean/AVE	Reliability/C. R.
System Convenience Satisfaction (FSCS)	FSCS1	The flipped learning screen is configured to be suitable for learning	3.960/0.550	3.90/0.50	0.745/0.830
	FSCS2	The content of the lecture to be taken can be easily searched	3.750/0.699		
	FSCS3	Download and streaming Speeds are appropriate	3.740/0.520		
	FSCS4	Contents are well uploaded to facilitate learning	4.040/0.596		
	FSCS5	Easy to access flipped learning contents	4.030/0.723		
Interaction Satisfaction (FIS)	FIS1	Notices, learning progression memos help learning	4.020/0.582	4.11/0.44	0.718/0.780
	FIS2	Questions and answers are exchanged fast and sincerely	4.260/0.635		
	FIS3	Communication between learner-teacher is smooth with band, e-mail, chatting, with band, e-mail, chatting, face to face hours	4.240/0.528		
	FIS4	Diverse and rich learning materials are shared among learners	3.630/0.572		
	FIS5	Flipped learning activities were helpful for learning	4.390/0.506		
Assistance Service Satisfaction (FASS)	FASS1	Operator's advance notices, and guides for learning are adequate	4.290/0.648	3.89/0.47	0.723/0.770
	FASS2	Diverse added services such as learning information are provided	3.630/0.582		
	FASS3	Diverse method for verification of learning effects are provided	3.580/0.587		
	FASS4	Assistance with contents for learning is appropriate	4.040/0.688		
	FASS1	Learning contents are composed to enable	4.400/0.543	4.11/0.44	0.683/0.790

(*continued*)

Table 2. (*continued*)

Variable	Division	Measurement variable	Mean/SD	Total mean/AVE	Reliability/C. R.
Learning Satisfaction (FLS)		accomplishing learning goals			
	FASS2	Video lectures are composed to induce learning motives and interest	3.870/0.559		
	FASS3	Lecture levels are appropriate for learning	4.000/0.624		
	FASS4	Learning content level is appropriate for learning	4.000/0.573		
	FASS5	Flipped learning was helpful for self-directed learning	4.280/0.506		

4 Conclusion and Suggestions

Based on the study results, implications for effective flipped learning can be presented as follows.

First, satisfaction with system convenience in flipped learning did not directly affect learning satisfaction. Given this, prior learning in out of classroom classes is judged to given burdens of learning to learners. As alternatives that can minimize the burden of prior learning, diverse methods should be sought such as the utilization of micro learning and provision of incentives for prior learning.

Second, satisfaction with interactions in flipped learning was identified as having direct effects on learning satisfaction. Given this, the interactive learning in classes in classroom enabled teachers' immediate and fast participation in learning that was not possible in traditional learning methods and the reconstructive learning intended to make the knowledge obtained by learners into senses of new information is judged to have affected learners' learning satisfaction. For proper learning guides for learners, teachers should be equipped with deeper teaching capabilities.

Third, satisfaction with assistance service in flipped learning did not directly affect learning satisfaction. Given this, in teaching-learning, other variables than the learner centered constructive learning variable had little effects on learning satisfaction.

Forth, satisfaction with system convenience in flipped learning did not have indirect effects on learning satisfaction mediated by satisfaction with assistance service. Given this, assistance related to learning information in out of classroom class is judged to have insignificant effects on learning satisfaction.

Fifth, satisfaction with interactions in flipped learning was identified as having indirect effects on learning satisfaction through mediation by assistance service satisfaction. Given this, when implementing learner centered constructive discussion and participation in classes in classroom, mediation by prior notices for learning, guide,

added service, evaluation methods, and support with diverse contents is judged to be a variable that can affect learning.

References

1. Bishop, J.L., Verleger, M.A.: The flipped classroom: a survey of research. In: ASEE National Conference Proceedings, Atlanta, GA, vol. 30, no. 9 (2013)
2. Deslauriers, L., Schelew, E., Wieman, C.: Improved learning in a large-enrollment physics class. Science **332**(6031), 862–864 (2011)
3. Bergmann, J., Sams, A.: Flip your classroom: reach every student in every class every day. International Society for Technology in Education, Washington, DC (2012)
4. A review of Flipped learning. http://www.flippedlearning.org/review
5. Lee, J.-Y., Park, S.H., Kang, H.-J., Park, S.-Y.: An exploratory study on educational significance and environment of flipped learning. J. Digit. Convergence **12**(9), 313–323 (2014)
6. Shon, J.G., Kim, B.W.: Design and implementation of content model for m-learning. Korea Inf. Process. Soc. **10**(4), 543–554 (2014)
7. Strayer, J.F.: How learning in an inverted classroom influences cooperation, innovation and task orientation. Learning Environ. Res. **15**(2), 171–193 (2012)
8. Baker, J.W.: The "Classroom Flip": using web course management tools to become the guide by the side. In: Selected Papers from the International Conference OFN College Teaching and Learning, vol. 11, pp. 9–18 (2000)
9. Flipped Learning Network. http://flippedlearning.org/site/default.aspx?PageID=1
10. Kim, M.K., Kim, S.M., Khera, O., Getman, J.: The experience of three flipped classrooms in an urban university: an exploration of design principles. Internet High. Educ. **22**, 37–50 (2014)
11. Wang, S.-K., Reeves, T.C.: The effects of a web-based learning environment on student motivation in a high school earth science course. Education Tech. Research Dev. **54**(6), 597–621 (2007)
12. Young-bae, K.: Development of the design principles on flipped learning support system. The Graduate School, Pusan National University, Department of Multimedia (2015)
13. Park, G.-H.: (The) Influence of College Educational Service Quality on Academic Motivation and Satisfaction: Focused on Comparison of Online/Offline Educational Services, The Graduate School, Keimyung University, Department of public Administration (2011)
14. Kim, Y., Shon, J.G.: A study on design of K-12 e-Learning system for utilization smartphone. Korean Soc. Internet Inf. **12**(4), 135–143 (2011)
15. Cho, A., Roh, S.-Z.: The analysis of structural relationships among self-directed learning ability, learning flow, learning attitude, academic satisfaction and achievement in cyber university. J. Educ. Technol. **12**(4), 849–879 (2013)

Development of UI Guideline for Senior Citizens' e-Learning Content

Myung In Kim and Yong Kim(✉)

Department of e-Learning, Graduate School,
Korea National Open University, Seoul, South Korea
dcesonia@gmail.com, dragonknou@knou.ac.kr

Abstract. The speed of aging increases day by day leading to changes in many shapes of society. In particular, the field of lifelong education out of the field of education has been expanded to diverse ranges and e-learning, which is a field of lifelong education and recognized as an alternative, is becoming gradually more important. In this situation, e-learning guidelines for silver generations are an essential element of the development of contents. The purpose of the present paper is to develop guidelines for e-learning UI designs for silver generation with verified validity. This author expects that e-learning sites applied with guidelines designed based on the present study will become more useful to silver generations and utilized by them.

Keywords: e-Learning · User interface · Content design · Senior citizen

1 Introduction

Information oriented society that has been developing rapidly led to intensive development of diverse technologies and thanks to the development of medicine, the ratios of aged populations have been rapidly increasing throughout the world. According to definitions by the UN, a society of a country with the ratio of aged populations exceeding 7 % of the population is an 'aging society' and a society of a country with the ratio of aged populations exceeding 14 % is an 'aged society.' [1]. South Korea has been also show high aging speeds. South Korea has become an 'aging society' in 2000 as the ratio of populations aged 65 years or more to the total population exceeded 7.2 % and is expected to become an 'aged society' in 2018 as the ratio of populations aged 65 years or more to the total population will be 14.3 % and a 'super aged society' in 2026 as the ratio will be 20.8 % [2]. One of important issues in aging society as such is the activation of lifelong learning for enhancement of the quality of life in the times of 100 years of human life. Because of physical characteristics of humans, as aging progresses, the aging of vision and tactile sense progresses particularly faster than other organs. These declines in visual cognitive functions of silver generations lead to the necessity to derive forms that can be more easily and clearly perceived and apply the forms to UI designs for e-learning. In addition, the declines in cognitive ability result in slower reactions to certain stimuli and remarkable declines in understanding and memory. That is, e-learning UI designs that can not only solve problems in contents but also enhance users' accessibility to and understanding of devices are necessary. To derive UI design guidelines for silver

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_116

generations, the present study began with research into literature regarding web UI design guidelines and e-learning UI designs already prepared and examining silver generations and e-learning environments. Based on this preceding study, a basic frame of e-learning UI design guides for silver generations was designed and the validity of the design was secured through an expert group. With the perception of the necessity of e-learning as an alternative for relief of information gap of silver generations and for their lifelong learning in aging society, which is a global phenomenon, the purpose of the present study was to present e-learning UI design guidelines to improve silver generations' accessibility to contents. This paper also hopes that the results of present study can be used as basic data for diverse contents for silver generations.

2 Related Works

Social welfare related laws of South Korea define elderly persons as people aged at least 65 years and the UN defines elderly persons as 'those who are 65 years old or older' based on time elapse units. 'Silver generation' is a phrase referring to the elderly otherwise that was contained in the 2003 'new words' source book of The National Institute of the Korean Language. Although population increase rates have been gradually decreasing throughout the world from 2000, silver population increase rates have been showing a tendency opposite to entire population increase rates.

When aging has progressed, declines in general functions appear including senses, perception skills, physical functions, and even psychological functions. However, the rates of use of the Internet among the elderly have been continuously increasing; to review the rates by age groups, the rate among the elderly aged 50–59 years in 2015 was 81.4 % (9.6 %p↑ compared to the previous year) and that among the elderly aged 60 years was 29.4 % (1.9 %p↑compared to the previous year) [3]. In addition, silver generations' desires for the quality of life has been switched from those centered on 'physical wealth' to those centered on 'qualitative value' indicating changes in population and social paradigms such as the switch to knowledge and information society and silver generations' lifelong learning courses will increase further.

UI design is ultimately the process to systematically structuralize the moments and processes of communication between artifacts and users, and in terms of actual work, UI design was defined as work to find rules of the methods of interactions between computer systems and users and make detailed designs for the elements [4].

The area for solving problems in web use for elderly users began to be studied in the 1990s and web interface design principles for elderly users considering body motion control, perception, recognition, and emotional aspects were studied in the USA [5].

In other studies, guidelines for the elderly presented in individual questions were extracted and were classified by adding programming, which is recommended in web standards, to interface items. The classified elements were reclassified into detailed design elements and check lists and recommended values for individual items were organized to prepare web interface design guidelines for the aged [6].

Since the most important item in e-learning is contents, UI designs made understanding and considering the characteristics, limitations, and potentials of e-learning for silver generations are necessary. Therefore, e-learning UI screens were studied based

on learners' demands and there is a study on the components of the interface of e-learning contents screens [7]. In addition, a study was conducted that divided e-learning UI design elements to consider even non-elderly persons [8].

3 UI Guidelines Design

3.1 Study Method and Procedure

The study method and procedures used to accomplish the purpose of the study were as follows. First, the largest items of e-learning guidelines for silver generations were composed based on the characteristics of silver generations, web UI design guides, and e-learning UI design guidelines identified through literature research and previous study reviews. Second, the content validity of the item "learning information", which is the most closely related with contents among the derived UI design guidelines items were verified by two educational engineering specialists and four design experts to present the components of detailed guidelines. Third, questionnaire surveys on the guidelines verified by experts were conducted with experts with at least three year careers in the designing of e-learning contents and the levels of importance and difficulties in use were identified through descriptive statistics to secure the validity and reliability of the guidelines.

3.2 UI Guidelines

The items of e-learning UI design guides for silver generations comprised learning windows, learning information, learning control devices, and identity and among them, the item learning information was constructed in detail. The constructed item is as shown in <Table 1> below.

Table 1. UI design guidelines

Construction units	Item
Character	Unity of fonts throughout the learning content – Use Gothic fonts for letters not larger than 13 pt and Ming-style fonts for letters not smaller than 15 pt – Avoid italic letters and underlined letters – Use both capital letters and small letters where applicable (English)
	Consistency of font sizes and thicknesses – Use 13 ~ 14 pt for body texts – Use 18 ~ 24 pt for title texts
	Clarity of division such as font spacing and line spacing – Avoid multi-column layouts or frame layouts and place 40 ~ 60 letters per line – Use 160 ~ 180 % lines of writing/text length per line should be 7 ~ 8 words

(*continued*)

Table 1. (*continued*)

Construction units	Item
	– Body texts shall be left-aligned(button: centering) – Font spacing shall be made appropriately based on font styles
Graphic	Image adequacy and definition – The resolution of related images should be maintained at least 96 ppi – Use images related to the lecture
	Consider the colors and sizes of the background and individual images – Differences in colors and brightness of the background and individual images are necessary – Margins are necessary outside each image (At least 5 pt)
	Maintain sizes and spaces in relation to surrounding elements such as texts – Provide text size adjusting function – Emphasize important texts by enlarging them or with bold effects
	Image shape – Maintain original shapes of images – Utilize illustrations and photos (Silver generation can understand more realistic contents more clearly)
Image	Clarity of lectures or lecture programs – Refrain from using difficult vocabularies and use terms without polysemy – Make the time of educational contents not to exceed 20 min per piece – In cases where there is a teacher, take chroma-keys and synthesize with the lecture images – Teacher hiding (support viewing only lecture programs) button – Use dark grey close to black for lecture program texts on bright color backgrounds
	Size and position of lecture screen – Position at the center of the lecture screen – Support lecture screen ratios in a range of 70 ~ 100 % of the entire monitor
	Lecture screen zooming and original screen size changing – Lecture screen zooming (Provide the function with a limit of 3 times at the maximum (within 200 %) and support returning to the original size)
	Support viewing the entire screen – Support viewing the entire lecture of lecture program
Voice	Constant pitches without any noise – Constant sound volume without any noise
	Adequacy of basic voice volume – Avoid high frequencies exceeding 4,000 Hz – Provide basic voices at low frequencies (1,000 ~ 2,000 Hz)
	Adequate sizes of subtitles when voices cannot be easily heard – Provide subtitles when voices cannot be easily heard(hiding/showing) – Subtitles should be at least 12 pt in size

3.3 Validity Verification

For the item "learning information" of which the sub items were determined in the present paper, expert opinions were collected for the second time through a questionnaire survey conducted with 30 hands-on workers who were currently designing e-learning contents at a school or an enterprise in relation to e-learning with at least three years of career. The collected data were analyzed for descriptive statistics using SPSS v. 21.

Three unit elements of characters in the area of learning information were presented. According to the results of analysis, 'Consistency of font sizes and thicknesses' (mean 4.43) showed the highest level of importance followed by 'Unity of fonts throughout the learning content' (4.33) and 'Clarity of division such as font spacing and line spacing (4.43).' The highest level of difficulties in use was shown by 'Clarity of division such as font spacing and line spacing (2.70)'.

Four unit elements of graphics in the area of learning information were presented. According to the results of analysis, 'Image adequacy and definition' showed the highest level of importance with a mean value of 4.53 followed by 'Consider the colors and sizes of the background and individual images (4.37)', 'Maintain sizes and spaces in relation to surrounding elements such as texts (4.23)', and 'Maintain original shapes of images (4.17)'. The highest level of difficulties in use was shown by 'Consider the colors and sizes of the background and individual images (2.80)'.

Four unit elements of images in the area of learning information were presented. According to the results of analysis, 'Clarity of lectures or lecture programs (mean 4.53)'showed the highest level of importance followed by 'Size and position of lecture screen (4.37)', 'Lecture screen zooming and original screen size changing (4.07)', and 'Support viewing the entire screen (4.03)'. The highest level of difficulties in use was shown by 'Lecture screen zooming and original screen size changing (3.53).'

Three unit elements of voices in the area of learning information were presented. According to the results of analysis, 'Constant pitches without any noise (mean 4.47)' showed the highest level of importance followed by 'Adequacy of basic voice volume (4.30)' and 'Adequate sizes of subtitles when voices cannot be easily heard (4.07).' The highest level of difficulties in use was shown by 'Adequacy of basic voice volume (2.73).'

4 Conclusion and Suggestions

In the situation of rapid aging, the e-learning contents market for silver generations has been quantitative growing and has been also established as an alternative for lifelong education that cannot be maintained with only traditional learning methods and forms. Therefore, e-learning interface designs for silver generations can be a factor that affects silver generations' satisfaction with e-learning contents when they study in e-learning environments. Therefore, in the present study, UI guidelines for silver generations were prepared first by referring to previous studies, analyzing related literature to organize the contents, and considering the results of studies and opinions of paper authors and UI guidelines for silver generations were designed after consulting expert groups with

the prepared items. The designed guidelines were classified into four items; learning windows, learning information, learning control devices, and identity and among them, detailed guidelines were designed for the item learning information, which is the most closely related with contents. Although the present study designed guidelines for UI design for silver generations through literature research and consulting with expert groups conducted a questionnaire survey with 34 learners, a small number of respondents, it is thought to have meaningful results in its own way in that it produced the results of analysis of items that are considered important in current silver generations' e-learning. In addition, if the developed e-learning UI guides are applied to the item learning information for production of e-learning contents for silver generations based on the results of the questionnaire survey, the efficiency of the site can be improved and learners' satisfaction can be enhanced to help the achievement of learning.

In follow-up studies, sub items should be determined for the individual items determined as large items and the validity of the sub items should be checked. Furthermore, in addition to the weakening of physical functions addressed in the present paper, studies that consider the psychological areas or aesthetic areas of silver generations should be conducted. If design guide items recognizing the characteristics of genders are constructed, another approach to the development of customized education type e-learning contents for silver generations can be developed.

References

1. U.N. Department of Economic and Social Affair: Toward a System of Social and Demographic Statistics, p. 178. United Nations, New York (1995)
2. Korea Network Information Center: Analysis of Current internet the middle and senior-aged (2006)
3. Ministry of science, ICT and Future Planning: The Survey on the digital divide index and status 2014 (2015)
4. Jin. J.A.: Analysis on Interest Factors for User Interface Design of Digital Textbook: Focused on the Relationship with Learning Motivation, Graduate School, Hongik University, Department of Film, Video and Animation (2012)
5. Lee, K.J.: Design Principle Development of Universal Interface on Educational Website Pages for the Elderly , The Graduate School, Pusan National University (2011)
6. Lee, H.J., Woo, S.H., Park, E.Y., Baek, S.C.: A study on the readability of web interface for the elderly user :focused on readability of Typeface. J. Korean Soc. Des. Sci. **20**(3), 315–324 (2007)
7. Choi,. B., Jung, J.: Analysis on e-Learning contents UI layout and research on effective UI layout of a screen demanded by learner. J. Korean Archit. **9**(2) (2009)
8. Lee, M.Y.: A study on e-learning UI design for the elderly based on universal design, The Graduate School, DongGuk University, Department of Multimedia (2013)

Full Duplex Relaying with Buffer Based on Cognitive Radio Technique

Junsu Kim, Doo-Hee Jung, Jeho Lee, and Su Min Kim(✉)

Department of Electronics Engineering, Korea Polytechnic University, Siheung, Gyeonggi-do 15073, Republic of Korea
suminkim@kpu.ac.kr

Abstract. In this paper, we propose a cognitive radio based power control scheme for full duplex relaying (FDR) system with buffer. The proposed scheme maximizes the system throughput by adaptively controlling the transmit power of relay by using underlay cognitive radio (CR) technique. We verify the performance of the proposed system with intensive simulations.

Keywords: Full duplex relaying · Buffer · Cognitive radio · Power control

1 Introduction

Recently, full duplex relaying (FDR) technique has gained a lot of interest from many researchers since its potential possibility for improvement of spectral efficiency over the conventional half duplex relaying. The main hurdle for FDR system is a self-interference problem which is induced because transmission and reception occur simultaneously in the same system. Hence, there have been various studies on the impact of the self-interference and how to mitigate it [1–3].

Meanwhile, in order to increase spectrum utilization efficiency, cognitive radio (CR) techniques have been studied widely in the viewpoint of spectrum sharing [4]. Although the purposes of the above two techniques are different, it is common that both techniques are the problems of handling interference. Hence, if we merge the two techniques, it would be possible to achieve the better performance than the existing approach. Among the various CR techniques, underlay CR technique is appropriate to adapt to FDR system since its simplicity and good performance [5].

In this paper, we propose CR based power control scheme for FDR system with buffer to maximize the system throughput by adaptively controlling the transmit power of relay. Moreover, we verified the performance of the proposed system with intensive simulations.

2 Problem Statement

Figure 1 depicts the system model considered in this paper with single source, single relay, and single destination. It should be noted that the relay in the system model has a first-in first-out (FIFO) traffic queue which can store the received data. In our system

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_117

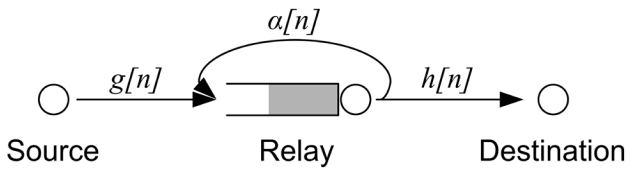


Fig. 1. System model

model, the relay works in full duplex mode, in which the relay transmits and receives simultaneously. Hence, unlike the usual half duplex relaying, the source and the relay can transmit at every time slot.

In the figure, channel gains $g[n]$ and $h[n]$ denote instantaneous complex channel gains of the source-relay and the relay-destination links, respectively, at time slot n. Channel gain $\alpha[n]$ represents self-interference component which is induced due to FDR. Note that we assume independent Rayleigh fading channels and $g[n]$, $h[n]$, and $\alpha[n]$ include both small and large scale fading components such as $|g[n]|^2 \sim Exp(\rho_{sr})$, $|h[n]|^2 \sim Exp(\rho_{rd})$, and $|\alpha[n]|^2 \sim Exp(\rho_{rr})$, where $Exp(\rho)$ denotes the exponential distribution with mean ρ, and ρ_{sr}, ρ_{rd}, and ρ_{rr} represent large scale fading of the source-relay, relay-destination and relay-relay links, respectively.

The source transmits in every time slot with variable rate according to the channel condition and the relay stores the arrived data in the queue and forwards the stored data from the queue simultaneously, i.e. full duplex operation. Infinite queue length is assumed as the first stage of study.

Let $P_S[n]$ and $P_R[n]$ denote the transmit powers of source and relay at time slot n and P_N represents noise power. Then, we define $\rho_S[n] = \frac{P_S[n]}{P_N}$ and $\rho_R[n] = \frac{P_R[n]}{P_N}$ as the normalized transmission powers of source and relay at time slot n. Main objective of this study is to maximize the system throughput by controlling the transmission powers. Specifically, we set the transmit power of the source as constant such as $P_S[n] = P_S$ and adjust the transmit power of the relay $\rho_R[n]$.

3 Proposed CR-Based Power Control Scheme for Full Duplex Relaying

3.1 System Throughput

Let $\gamma_R[n]$ and $\gamma_D[n]$ denote the received signal-to-interference and noise ratio (SINR) of the relay and the signal-to-noise ratio (SNR) of the destination at time slot n, respectively. Then, they can be shown as

$$\gamma_R[n] = \frac{|g[n]|^2 P_S}{P_N + |\alpha[n]|^2 P_R[n]} = \frac{|g[n]|^2 \rho_S}{1 + |\alpha[n]|^2 \rho_R[n]}, \tag{1}$$

$$\gamma_D[n] = |h[n]|^2 \frac{P_R[n]}{P_N} = |h[n]|^2 \rho_R[n]. \tag{2}$$

Hence, at time slot n, the traffic amount that the source can transmit is limited by (1) and the traffic amount that the relay can forward is determined by (2). If we denote traffic arrival and departure of the relay as $A[n]$ and $D[n]$, then they can be formulated as

$$A[n] = \tau \log_2(1 + \gamma_R[n]), \tag{3}$$

$$D[n] = \tau \log_2(1 + \gamma_D[n]), \tag{4}$$

where τ is a time slot length in sec. Please note the unit of (3) and (4) is bits/Hz. Since the relay has a queue with infinite length, data can be left in the queue. Let the queue length at time slot n be $Q[n]$. Then, it can be written as

$$Q[n] = Q[n-1] + A[n] - \min\{Q[n-1], D[n]\}, \tag{5}$$

and $Q[0] = 0$ bits/Hz. Finally, the traffic rate delivered to the destination, which is the system throughput $R[n]$ in bps/Hz, is formulated as

$$R[n] = \frac{1}{\tau} \min\{Q[n-1], D[n]\}. \tag{6}$$

Therefore, ergodic system throughput can be obtained as

$$\overline{R} = \frac{1}{\tau} E\{\min\{Q[n-1], D[n]\}\}, \tag{7}$$

where $E\{\cdot\}$ denotes expectation.

If the relay raises the transmit power $\rho_R[n]$ to achieve the higher departure rate, it induces the higher self-interference, which is $|\alpha[n]|^2 \rho_R[n]$ term in (1), then the arrival rate in (3) decreases which yields short queue length as in (5).

3.2 Proposed CR-Based Power Control Scheme

As explained above, the arrival rate and the departure rate are in tradeoff relation because of the self-interference channel which is induced by FDR. In order to maximize the system throughput, we should minimize the self-interference while maximizing the departure rate by adjusting the transmit power of relay adaptively. To achieve this goal, we introduce underlay CR technique to control the transmit power of relay.

Underlay CR system allows simultaneous transmission of secondary transmitter while primary system is active if the interference to the primary system is kept under a certain level, which is called interference temperature [4]. To apply this technique to our system, we regard the source to relay receiver as the primary system and the relay transmitter to destination as the secondary system. Hence, the relay transmitter should

adjust its transmit power such that the interference to the relay receiver is less than the interference temperature.

If we let I denote the interference temperature, then the transmit power of relay at time slot n should meet the following condition.

$$|\alpha[n]|^2 \rho_R[n] \leq I \tag{8}$$

Then, we are able to control the transmit power of relay according to the following rule.

$$\rho_R[n] = \min\left\{\frac{I}{|\alpha[n]|^2}, \rho_S\right\} \tag{9}$$

It is worth to note that setting the maximum power of the relay as ρ_S, which is the constant transmit power of source, is reasonable for total power management of the system.

4 Performance Evaluation

4.1 Comparison with Conventional Full Duplex Relaying

Figure 2 compares the system throughput of the proposed scheme with that of the conventional FDR system. Note that the conventional FDR system does not adjust relay power but uses a constant power as $\rho_R[n] = \rho_S$. Figure 2 shows the case when all average channel qualities are even such that $\rho_{sr} = \rho_{rd} = \rho_{rr} = 0$ dB. As shown in the figure, when ρ_S increases the proposed scheme with appropriate values of I outperforms

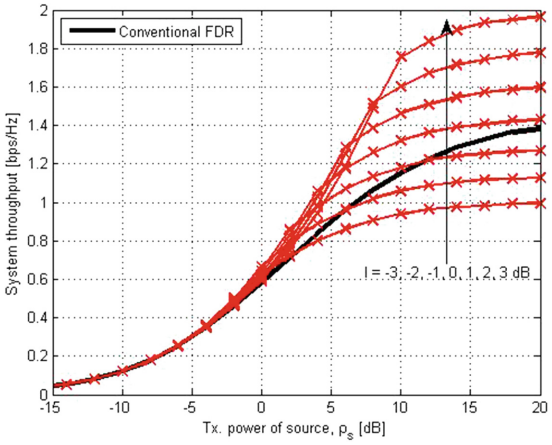


Fig. 2. Throughput comparison with conventional system, $\rho_{sr} = \rho_{rd} = \rho_{rr} = 0$ dB, $I = -3, -2, -1, 0, 1, 2, 3$ dB

the conventional scheme. Especially, it can be seen that as more self-interference is allowed, higher throughput can be achieved.

4.2 Impact of Interference Temperature

One important parameter is the interference temperature. As shown in the previous results, if we use higher interference temperature, we can also achieve higher system throughput. Is this always true? In this subsection, we evaluate the impact of interference temperature on the system throughput.

Figure 3 investigates the arrival rate, departure rate, and system throughput according to the varying interference temperature when $\rho_{sr} = \rho_{rd} = \rho_{rr} = 0$ dB, which is the same scenario as in Fig. 2, and $\rho_S = 10$ dB. As I increases, i.e. more self-interference is allowed, the arrival rate decreases while the departure rate increases. When the arrival rate is greater than the departure rate, the queue length will become longer and, in the opposite case, the queue will be empty. Hence, as shown in Fig. 3, maximum throughput can be achieved when the arrival rate and the departure rate are balanced. In other words, there exists an optimal interference temperature which maximizes the system throughput. The optimal value of I is determined when the arrival rate and the departure rate are balanced.

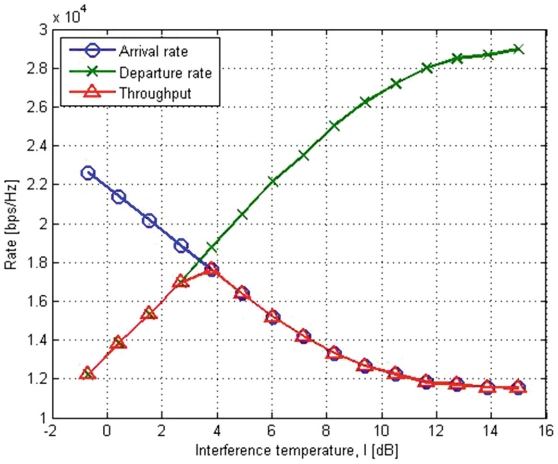


Fig. 3. Impact of interference temperature, $\rho_{sr} = \rho_{rd} = \rho_{rr} = 0$ dB and $\rho_S = 10$ dB

5 Conclusions

In this paper, we proposed cognitive radio based power control scheme for FDR system with buffer. The proposed scheme maximizes the system throughput by adaptively controlling the transmit power of relay by using underlay cognitive radio technique. We verified the performance of the proposed system with intensive simulations.

Through the simulation results, we showed our proposed scheme outperforms the conventional FDR system in wide region. We also investigated the impact of interference temperature on the system throughput to show that there exists an optimal interference temperature value which maximizes the system throughput.

In the future research, it is needed to formulate ergodic system throughput in general case to obtain optimal throughput and corresponding interference temperature. Moreover, the study considering the limited queue length is also required.

Acknowledgments. This work (Grants No. C0396777) was supported by Business for Cooperative R&D between Industry, Academy, and Research Institute funded Korea Small and Medium Business Administration in 2016.

References

1. Kang, Y.Y., Kwak, B.J., Cho, J.H.: An optimal full-duplex af relay for joint analog and digital domain self-interference cancellation. IEEE Trans. Commun. **62**(8), 2758–2772 (2014)
2. Kim, H., Lim, S., Wang, H., Hong, D.: Optimal power allocation and outage analysis for cognitive full duplex relay systems. IEEE Trans. Wirel. Commun. **11**(10), 3754–3765 (2012)
3. Zarbouti, D., Tsoulos, G., Athanasiadou, G.: Effects of antenna array characteristics on in-band full-duplex relays for broadband cellular communications. ICT Express **1**(3), 121–126 (2015)
4. Mitola, J., Maguire, G.Q.: Cognitive radio: making software radios more personal. IEEE Personal Commun. **6**(4), 13–18 (1999)
5. Jovicic, A., Viswanath, P.: Cognitive radio: an information-theoretic perspective. IEEE Trans. Inf. Theory **55**(9), 3945–3958 (2009)

Design of Docking Drone System Using P-PID Flight Controller

Beck Jong-Hwan, Pak Myeong-Suk, and Kim Sang-Hoon(✉)

Department of Electrical, Electronic and Control Engineering,
Hankyong National University, Anseong, Republic of Korea
vlwha0240@naver.com, nicems@nate.com, kimsh@hknu.ac.kr

Abstract. Recently, drones have been used widely because of low-cost, mobility and high utility, and it can be applied in many fields such as surveillance, guidance, military and hobby, etc. using docking modules. This paper proposes a control method for docking drones using P-PID controller. This drone is applied the docking method using marker detection, and it is made for several experiments for stable docking. In this paper, we also describe a P-PID controller for docking drones' flight. Experimental results have shown docking drones' considerable material and the proposed method is more efficient than single PID method.

Keywords: Drone · Docking module · Flight algorithm · Feedback controller · Image processing Marker detecting

1 Introduction

Until now, there have been many researches about UAV (Unmanned Aerial Vehicles) called 'drones' [1]. Quadcopters are complex nonstationary nonlinear systems, and tuning the controller gains usually is a challenging problem. Yet, load and parameter changes like inertia, and lateral and up-down wind disturbances can significantly affect the performance of the control system [2]. They are aircraft either controlled by pilots from the ground or autonomously following a preprogrammed mission. Especially, a quadcopter not only has the same merits as any other aircraft but also has low cost, high utility and no need to consider the influence of ground environment. Although maximum operating time is very short because of power consumption, it can be applied in many fields such as surveillance, guidance, military and hobbies, etc. For the purpose of extending the function of drone, docking is a promising approach. 'Docking' means 'joining of two vehicles'. In other word, it means not just 'adhesion' but 'combining two functions'. It is worth researching due to the point that docking modules can be useful for drones. Meanwhile, there have been many studies about how to stabilize drones' attitude control system for flight. It is difficult to control the attitude using only single PID controller, so many drone use each algorithm.

Basic quadcopter frame is needed to be built well in order to control the attitude [3]. The paper deals with flight stability by examining drones' material. Each material has its advantages and weakness. In addition, we propose a method to combine two robots, and the robot's flight using a control method for docking drones using P-PID

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_118

controller. P-PID controller consists of angle feedback loop (using P controller, outer loop) and angular velocity feedback loop (using PID controller, inner loop). Values gained through an experiment of attitude control are refined by Kalman filter and the outcome PID values are applied to the drone to complete the final design.

2 Docking Robot System

2.1 Outline

As shown in Fig. 1, the proposed drones' system consists of body, controller and embedded board. 'Docking drone' has two operating modes and changes its motion depend on each mode. The body is comprised of 4-leg served as landing gear and a grabber with quad-rotor. 4-leg body part includes MCU board and sensors and it has to endure weight of quad-rotor's part. The drone is made to be solid and light [4]. Similarly, drone part can lift other parts. In addition, we selected the controller to be appropriate for the mobile phone application with portability and affordable.

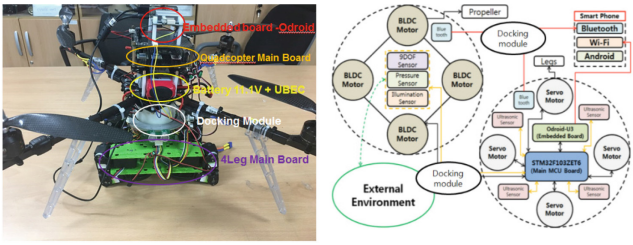


Fig. 1. This is the drone's appearance and system. Docking drone consists of embedded board, MCU, battery and docking module.

There are 4-leg for walking in 4-Leg part; it can avoid obstacles thanks to ultrasonic sensors. It is controlled by Android application a Bluetooth. If the robot starts to fly, the operator control legs and it can use grabber. Additionally, the system is included embedded board for image processing, so the user can retrieve real-time image, and understand surroundings.

2.2 Docking Method

As shown in Fig. 2, the docking method is very simple, and intuitive. The docking module consists of ABS Plastic gears and motor. If the quadcopter is located in marker area detected by image processing, the motor starts spinning and gear is working. The four plastic bars are fixed and docking is completed (Fig. 3).

Above figure is the flowchart of docking using image process. The camera located at quadcopter's underneath detect a marker such as above figure, and body move on the docking object. Lastly, the docking module is working as shown in Fig. 2. And docking process is complete.

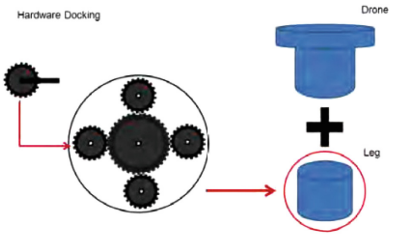


Fig. 2. Docking module. This is made of ABS Plastic, 190*190*79.5 (mm).

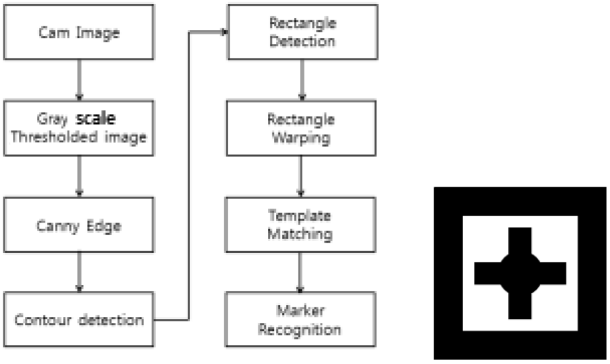


Fig. 3. The flowchart of docking using image processing and a marker for docking.

2.3 P-PID Control Algorithm

A Proportional-Integral-Derivative controller (PID controller) is a control loop feedback mechanism commonly used in industrial control systems [4]. A PID controller continuously calculates error value as the difference between a measured process variable and a desired set-point. The quadcopter is able to make a stable flight depending on its PID tuning parameter. The controller attempts to minimize the error over time by adjusting of a control variable, such as the position of a control valve, a damper, or the power supplied to a heating element, to a new value determined by a weighted sum [5]:

$$u(t) = K_p e(t) + K_i \int_0^t e(t)dt + K_d \frac{de(t)}{dt} \tag{1}$$

As shown in (1), P accounts for presenting values of the error, I accounts for past values of the error, and D accounts for possible future values of the error, based on its current rate of change [4]. We applied P-PID algorithm such as Fig. 4. There is no algorithm using only angular error, in fact the algorithm for quadcopters consists of inner loop and outer loop. The inner loop calculates angular error, and fixed quadcopter's Euler angle using the P controller, then the outer loop adjusts similarly gyro rate using the PID controller. Finally, calculated value input into ESC. All actual value

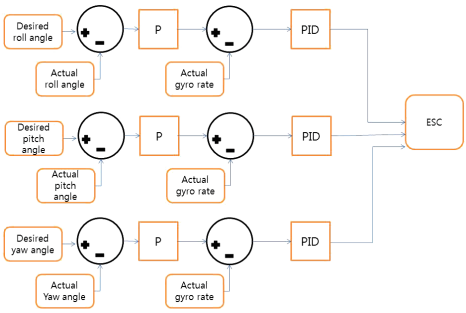


Fig. 4. P-PID Algorithm for attitude control

is measured by sensors. We select a 9-DOF sensor to measure actual angle and gyro rate. Used sensor transmits sensor's heading and attitude value to MCU.

3 Experiments

3.1 Durability Test

The docking body's durability is very important. If the body's durability is weak, there will be no stability in flight. So it must be considered body's material. The experimental results are summarized in Table 1. In this test, we consider weight of the four-leg part in several kinds of material.

The experiment was proceeded each material's body, paper, carbon and Aluminum. The stability value follows below equation.

$$\text{Stability} = \frac{\text{Accuracy value}}{\textit{Full} - \textit{flight time}} \tag{2}$$

Accuracy value means the number of successful landings in 100 tests. At first, the paper body is not efficient for control because of it durability. Aluminum body is lightweight but it is very stable for each tests.

Table 1. Durability test

Material	Total weight (kg)	Stability
Paper	1.9	Control disable
Carbon	2.03	0.52
Aluminum	2.79	0.83

3.2 P-PID Controller

As shown in Fig. 5, two bars attached four motors are crossed each other, and those are lifter. The lifter's movement make lift power and drone can fly.

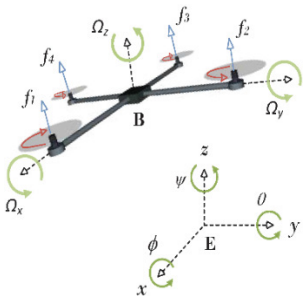


Fig. 5. Quadcopter coordinate system

Table 2. Drone model Parameter

Parameter	Value	Unit
J_x	0.7	Nms^2/rad
J_y	0.7	Nms^2/rad
J_z	1.1	Nms^2/rad
m	2.67	Kg
l	0.4	m
ρ	0.0046	

Table 2 shows the model drone's parameter, **ρ** are moments of inertia for each x, y, z-axis, m is the docking body's weight, l is the length from the center, also anti-torque proportional factor. The controller follows below equations.

$$u = Bf \tag{2}$$

$$B = \begin{bmatrix} 1/m & 1/m & 1/m & 1/m \\ 0 & l/J_x & 0 & -l/J_y \\ -l/J_y & 0 & l/J_y & 0 \\ \rho/J_z & -\rho/J_z & \rho/J_z & -\rho/J_z \end{bmatrix} \tag{3}$$

u means control input, f means quadcopters lift power. The quadcopter is able to start flying only if f is bigger than gravity force. So, we can derive control input u from Eqs. (2) and (3).

3.3 Result

Above graph is response curve for roll and yaw angle. the convergence time is about 3 ms in roll and yaw angle. The graph shows little oscillation in convergence time. (Fig. 6).

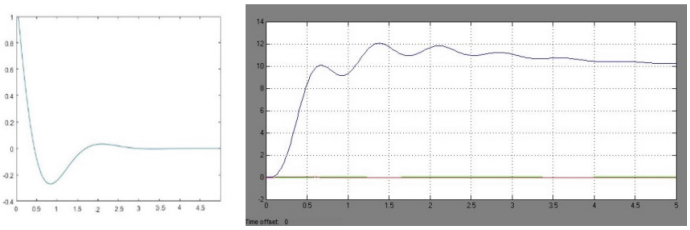


Fig. 6. Response curve for roll angle (left) and yaw angle (right)

4 Conclusion

As the quadcopters are getting more popular, modified mobile robot is getting more important and practical. This paper describes a drone model using docking module and P-PID controller for flight. The detailed control method for combining quadcopter and 4-leg robot discussed in this paper. We designed a quadcopter with high mobility, obstacle avoidance ability, and low-cost components. This quadcopter has four rotors and can fly stably using spin speed of rotors and P-PID control algorithm. As a result, we have shown the efficiency of selected materials and the algorithm. In addition, the drones could be useful using docking modules.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A01057518).

References

1. Saputra, O.D. Shin, S.Y.: Modeling and roll, pitch and yaw simulation of a quadrotor. Korea Institute of Communication Sciences, p. 21 (2013)
2. Domingos, D., Camargo, G., Gomide, F.: Autonomous fuzzy control and navigation of quadcopters. In: 4th IFAC Conference on Intelligent Control and Automation, vol. 49, no. 5, pp. 73–78 (2016)
3. Runqi, L., Lee, K.H., Nam, J.M., Nam, B.W., Tuepeng, W., Changkun, D.: A study based on mobile devices for quadcopter attitude control. Soc. Cadoam Eng., 141 (2015)
4. Beck, J.-H., Park, M.-S., Jeon, J.-S., Kim, S.-H.: Design of walking drone for surveillance. In: The 2016 World Congress on Information Technology Applications and Services (World IT 2016), world-it 54 (2016)
5. Araki, M.: PID Control. Control Syst. Robot. Autom. **2**
6. Lapp, M., Ahmadbeygi, S., Cohn, A., Tsimhoni, O.: A recursion-based approach to simulating airline schedule robustness. In: WSC 2008 Proceedings of the 40th Conference on Winter Simulation, pp. 2661–2667 (2008)

Lightweight Security for Underwater IoT

Sun-Ho Yeom, Jung-Il Namgung, Soo-Young Shin,
and Soo-Hyun Park(✉)

Department of Financial Information Security,
Kookmin University, Seoul, South Korea
junsan86@outlook.com, greenji21@naver.com,
{sy-shin,shpark21}@kookmin.ac.kr

Abstract. Becoming IoT era, RF communication on terrestrial network gets faster speed and wide communication coverage. However, because of RF has vulnerability on underwater environment, the underwater acoustic is widely adopting for replacing RF. As a result of efforts, we can suggest 'underwater IoT' model and its security challenge by listing UIoT security consideration and comparing underwater security protocols which are presented. Additional, we discuss side channel attack issues.

Keywords: UAC · UIoT · Security · SCA · CCM-UW · UW-AKE

1 Introduction

Underwater acoustic sensor network (UWASN) become the most widely suggested communication technique in underwater environment [1]. On the other hand, recently water resource utilization beyond the non-profit areas such as environmental monitoring, exploration of deep sea, becoming more popular in large-scale commercial business for examples "underwater data center" [2], "marine plant", and "sea farm". The utilization of information has changed 'vertical' to 'horizontal' using IoT/M2M. As a result, this will increase interconnectivity between different networks, but it also rises an issue of security. This paper describes the characteristics of Underwater Acoustic Communication (UAC) and define Underwater-IoT (UIoT). In this paper we compared the applicability of the underwater security protocols and explain the considerations of security in UIoT environment. The structure of the paper is as follows. Section 2 is related research which defines the characteristics of UAC and UIoT. Section 3 explains UIoT security consideration points, and compare underwater-based security protocols. In addition, we commented the lack of study on Side-Channel Attacks in Underwater Environment. In conclusions we mentioned the future research directions.

2 Underwater Internet of Things

This section defines the feature of UAC and UIoT.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_119

2.1 Underwater Acoustic Communication (UAC)

Radio Frequency is used most widely as a Wireless communication method in the terrestrial area. However, in underwater environment, it is almost impossible because of attenuation. Therefore, acoustic communication is used as an underwater communication method instead of RF [3]. The UAC must overcome the following restrictions to establish a complete communication. (a) Low bandwidth, (b) Long-latency, (c) High bit error, and (d) Multipath.

2.2 Definition of UIoT

(1) UIoT is the expansion of UWASN and defined as the communication network of underwater and provides interconnectivity between other networks, such as ground, and air.
(2) According to the latest trend, the data generated from the sensors can be operated to become a meaningful information by applying big data and semantic techniques. Figure 1 is the conceptual diagram of UIoT.

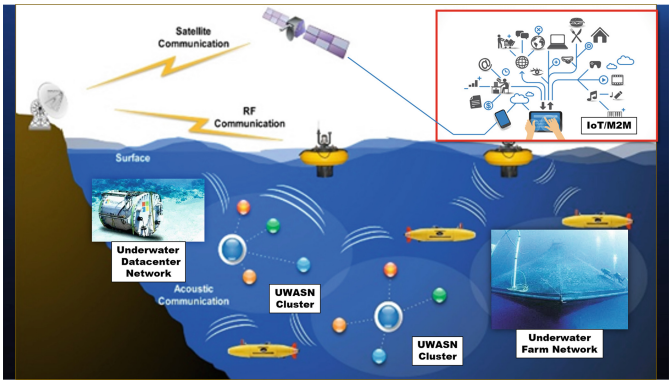


Fig. 1. Conceptual diagram of UIoT

3 UIoT Security

3.1 Consideration Points of UIoT Security

The following are the security consideration points for UIoT.

- Lightweight: Security with a larger bandwidth at UAC is too difficult, because of low transmission capability. It is essential to reduce the weight of entire payload, including the security protocol.
- Low-Power: Processor should be designed with less power consumption.
- Authentication & authorization: UIoT devices must be authenticated and authorized before interconnection.

- Intrusion Prevention: The underwater network used key exchange process for authentication. In this process it is exposed to three kinds of attacks such as 're-transmission', 'reflection', and 'mediators'. So it should have the intrusion prevention technology against them.
- Re-authentication: Due to the flow rate of underwater, the UIoT devices are re-positioned frequently than ground equipment. In that case authentication should be promptly discarded. Re-authentication should be established faster than the first time authentication process when node wants to interconnect to UIoT network.
- Data Encryption: Because of UAC features sustaining communication phase is not easy. Signal loss and connection discard is more frequently than Air RF. Therefor for intensifying security, adjust policy of data encryption it has big data size like AES, SEED even though DES is too hard to UAC network. Alternatively, should be considered lightweight data encryption algorithms and protocols.
- Side channel attack: Almost UIoT networks can be deployed in the distant sea. In this case, not only overcoming underwater communication problem but also concerning its battery life. For the saving power, remote administrator should be made difficult to trace networks location and mange devices status by real time. These vulnerabilities easy to make enemies to steal or hacking UIoT devices by the approaching network very closely and although it has powerful data encoding solutions that doesn't matter. This is risk of 'Side-Channel Attack'.

-Example of underwater SCA-

(1) Steal devices.
(2) Duplicating or hacking its chipsets.
(3) Analyze encoding algorithm and communication process.
(4) Make spy device for replacing stolen thing or hacking machine.
(5) Install '(4)' units.
(6) Even if the entire network is encrypted, it is easy to steal the original message and for the distorting UIoT network information, manipulate data or its communication process.

3.2 UIoT Security Protocols

This section analyzes the UIoT security protocols candidates.

- CCM-UW [3]: It is modified to adapt CCM*[1] operate mode for adjusting in UAC environment. Supports Step 6 security level, and provides a different energy consumption and security for each step. Deigned to the MAC[2] layer security. Low bandwidth and signal loss in the demonstration test result lacked to optimization.
- UW-AKE [4]: The lightweight authentication and key issuing protocol. Proven reliability between the members, and to share a secret key. A separate key exchange

[1] (Counter with CBC-MAC) It is based on the AES and ARIA algorithms.
[2] Media Access Control.

process by combining the authentication value to the authentication process is omitted to achieve a weight reduction. To reflect the re-authentication procedure according to the movement was always designed to maintain network security. It has not been validated in a lab environment.

- SeFLOOD [5]: It is underwater-based message exchange authentication protocol developed for defending vulnerabilities 'Flooding', 'Spoofing', 'DoS'. Although underwater security protocols, because it is designed for fixed and stable power supply performance device, it is difficult to view the appropriate security protocols for UIoT.

Table 1. The analysis is that they satisfy the above security protocols UIoT security conditions. The results are all set forth in the three protocol does not meet conditions. Since the UAC itself has a limited ability to communicate, it is necessary to describe in detail, but the scope of this restriction, fail to express the "qualitative". The experiment is also necessary that the intended work as much in the real world, because the rest of the proposed protocol, except for the CCM-UW are only a simulation.

Table 1. Analysis of underwater security protocols

Security condition	CCM-UW	UW-AKE	SeFLOOD
Lightweight	Different by 6 step security level	Combine authentication key	Not affordable for UIoT
Low-Power	Different by 6 step security level	Not consider	Not consider
Authentication & authorization	CBC-MAC[a]	Support	Support
Intrusion prevention	Frame counter by MAC	One time authentication (OTP)	Encrypted unicasting
Re-authentication	Not consider	Support but not efficient	Not consider
Data encryption	Different by security level	Combine en/decryption key	Conceptual

[a]Cipher Block Chaining-Message Authentication Code

4 Conclusions

In this paper, we explore the proposed security considerations applicable UIoT and security. As a result, the security considerations are suggested to stay at the level of 'qualitative', given the security protocols required a demonstration. On the other hand, is a side-channel attack research was inadequate compared to its severity. Therefore, the future research will be an experiment to analyze specifically those disclosed lightweight security protocols according to the numerical criteria. It will also issue by establishing the appropriate scenario presented in side-channel attack.

Acknowledgments. This research was supported by Department of Financial Information Security (BK21 + Future Financial Information Security Specialist Education Program Group), Kookmin University.

References

1. Akyildiz, I.F., Pompili, D., Melodia, T.: Underwater acoustic sensor networks: research challenges. Ad Hoc Netw. **3**(3), 257–279 (2005)
2. Microsoft is experimenting with underwater datacenters. http://natick.research.microsoft.com/
3. Ibragimov, M., Lee, J.-H., Kalyani, M., Namgung, J., Park, S.-H., Yi, O., Kim, C.H., Lim, Y.-K.: CCM-UW security modes for low-band underwater acoustic sensor networks. Wirel. Pers. Commun. 1–21 (2016)
4. Park, M., Kim, Y., Yi, O.: Light weight authentication and key establishment protocol for underwater acoustic sensor networks. J. Korean Inst. Commun. Sci. **39**, 360–369 (2014)
5. Dini, G., Duca, A.L.: SeFLOOD: a secure network discovery protocol for underwater acoustic networks. In: IEEE Symposium on Computers & Communications, pp. 636–638 (2011)

Image Based Video Querying Algorithm Using 3-Level Haar Wavelet Transform Features

Changseok Bae[1(✉)], Yuk Ying Chung[2], and Jeunwoo Lee[3]

[1] Department of Electronics, Information and Communications Engineering, Daejeon University, 62 Daehak Ro, Dong Gu, Daejeon 34520, Korea
csbae@dju.ac.kr
[2] School of Information Technologies, The University of Sydney, University of Sydney J12, Sydney, NSW 2006, Australia
vera.chung@sydney.edu.au
[3] SW and Contents Research Laboratory, ETRI 218 Gajeong Ro, Yuseong Gu, Daejeon 34129, Korea
ljwoo@etri.re.kr

Abstract. Surveillance cameras and smartphone cameras produce huge amount of video data in our daily life. Effective utilization of huge video data is emerging as a new big problem in the ubiquitous video intelligent systems. Searching for a specific frame in video stream is one of challenging issues in this area. This paper proposes an image based video querying algorithm using 3-level Haar wavelet transform features. Hierarchical decomposition of wavelet transform enables to use features in both space and scaling domains. This paper employs 3-level Haar wavelet feature of an image to query matched frame in a video stream. In experimental results, we can find that the proposed algorithm shows about 1–8% better performance in accuracy than other algorithms.

Keywords: Video query · Haar wavelet transform Video intelligent system

1 Introduction

It is very common to encounter 24-h recording surveillance cameras in both public and private sectors of our living space. In addition, we can record videos of special events very easily using our smartphones. Modern ubiquitous imaging devices produce huge amount of video data in every moment. Recording video is very easy and common in these days, however effective utilization of huge video data is a big problem. Especially, searching for a specific frame in video stream is one of challenging issues for using video data effectively [1].

Figure 1 illustrates general structure of image based video querying system. The video querying system consists of 3 components such as feature extraction, database clustering, and matching/searching.

This paper proposes a robust image based video querying algorithm using 3-level Haar wavelet transform features [2]. Wavelet transform decomposes input image hierarchically. Thus, wavelet transform enables to use features in both space and scaling

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_120

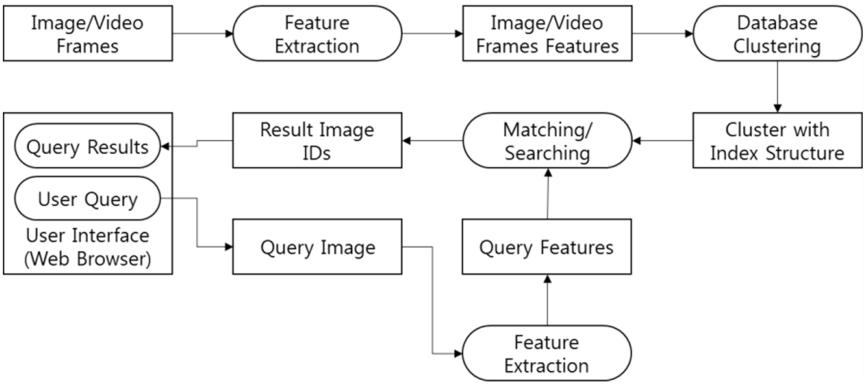


Fig. 1. General structure of image based video querying system.

domain. This paper considers 2 most representative wavelet transform methods such as Haar wavelet and Daubechies wavelet [3].

This paper consists as follows. Section 2 briefly describes related works, and Sect. 3 explains the proposed scheme. Then, in Sect. 4, we present experimental results. Finally, we provide conclusions and future works.

2 Related Works

This paper considers features of 2 wavelet transform methods and general RGB histogram for performance evaluation. This section explains these 3 feature extraction schemes very briefly.

The RGB histogram model is usually used in computer monitors and graphics [4]. The mixture of colors (Red, Green and Blue) are used to specify a particular color on the screen. Along each axis of the color cube the colors range from no contribution of that component to a fully saturated color. Any point within the cube is specified by three numbers, namely, R, G, and B triple. The diagonal line of the cube from black (0, 0, 0) to white (1, 1, 1) represents all the greys, that is, all the red, green, and blue components are the same [4].

The Haar wavelet [2] is known as the simplest wavelet, and it is defined as:

$$\psi_{jk}(x) \equiv \psi(2^j x - k)$$

for a nonnegative value j and $0 \leq k \leq 2^j - 1$. Here, $\psi(x)$ is defined as:

$$\psi(x) \equiv \begin{cases} 1, & 0 \leq x < 1/2, \\ -1, & 1/2 \leq x < 1, \\ 0, & otherwise. \end{cases}$$

The Daubechies wavelet [3] is named after its inventor, a mathematician Ingrid Daubechies. The Daubechies D4 transform has four wavelet and scaling function coefficients. The scaling function coefficients are:

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}},$$
$$h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}.$$

Here, the wavelet function coefficient values are: $g_0 = h_3$, $g_1 = -h_2$, $g_2 = h_1$, and $g_3 = -h_0$. The scaling and wavelet functions are calculated by taking the inner product of the coefficient and four data values. The equations are as follow:

Daubechies D4 scaling function:

$$a_i = h_0 s_{2i} + h_1 s_{2i+1} + h_2 s_{2i+2} + h_3 s_{2i+3}$$

Daubechies D4 wavelet function:

$$c_i = g_0 s_{2i} + g_1 s_{2i+1} + g_2 s_{2i+2} + g_3 s_{2i+3}$$

3 Wavelet Feature Extraction

Prior to the wavelet transform, the input image has to be resized so that the size of images is quadratic in both the database and the query image. The resizing can be achieved by cutting an image, or by averaging several pixels to obtain a single pixel to decrease the image size. For instance, a 160 by 160 image is to be resized to 128 by 128. As for a better descriptor in color definition, the image is transformed from RGB color space to YCbCr, which is more appropriate for the wavelet transform. Once the image is resized and the color space is converted into YCbCr color space, the wavelet transform can be applied.

In order to perform 3-level wavelet transform (both Haar wavelet and Daubechies D4 wavelet), we firstly separate an image into different frequency bands in 3 hierarchical levels, as shown on Fig. 2.

The feature vector generated by wavelet transform has 28 dimensions and it was formed by 2 color components of an image, such as color distribution and color variation. The first 4 elements of the feature vector D0 to D3 are the color variation (mean deviation) in Y color channel in the LL3, HL3, HH3 and LH3 bands. The mean deviation is the mean of the absolute deviations of a set of data. Feature vectors of wavelet are described as:

(D0, D1, D2, D3, ← Color variation components
H0, H1, H2, H3, H4, H5, H6, H7,
H8, H9, H10, H11, H12, H13, H14, H15, ← Color histogram components
H16, H17, H18, H19, H20, H21, H22, H23)

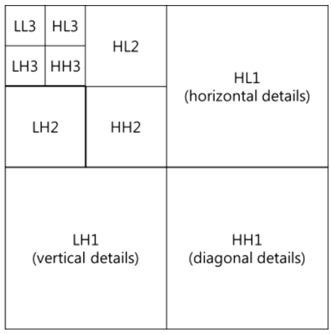


Fig. 2. 3-level wavelet transform.

The D0 to D3 compose the color variation component of the feature vector. These components represent the "color changes" of an image. For instance, an image contains similar or same color over different pixels will have low color variation ratio, while an image contains sharp and frequent color changes will have a higher color variation ratio. The color variation component is considered as the least description of an image.

The H0 to H23 compose the color histogram component of the feature vector, and they are the histograms in the LL3 band in three color-channel; H8 to H15 are the histograms in the Cb color channel and H16 to H23 are histograms in the Cr color channel. The histogram information consists of more detailed description of an image and it describes the color distribution of an image.

In comparison to the feature vectors that only consist of color histograms, the feature vectors produced by wavelet capture both of the color changes and color histograms information, which produces a more detailed description of an image. Thus, these values describe images very accurately.

2-level wavelet transform is similar to the 3-level wavelet transform, except that it only apply 2-level transform to an image. Similar to 3 level transform, 2-level transform has 28 dimensions and it was formed by 2 color components of an image, they are color distribution and color variation. The first 4 elements of the feature vector D0 to D3 are the color variation in Y color channel in the LL2, HL2, LH2 and HH2 bands. The H0 to H23 are the color histograms in three color channels in the LL2 band.

4 Experimental Results

The purpose of the experiment is to examine both the querying performance and accuracy in RGB histogram (Color-based) and Wavelet transforms (Wavelet-based) with particular focus on video contents. As illustrated previously, we have compared 5 algorithms such as:

- RGB Histogram
- Haar Wavelet (2 Level transform)
- Haar Wavelet (3 Level transform)

- Daubechies Wavelet (2 Level transform)
- Daubechies Wavelet (3 Level transform)

Each of the above algorithms was matched with clustering techniques such as Forgy method [5], K-Means method [5], and Isodata method [6].

Total of 20 query images were randomly selected from the sequence image frames among the 24 video streams, and they were excluded from the image database for a more accurate experimental results. For a given query image, there will be a specific number of top images returned by the system. The accuracy measurement is defined as:

$$\text{Accuracy} = \text{C}/\text{T},$$

where C is the number of images returned from the class of which the query image exists and T is the total number of images.

This is done to examine the performance of different algorithms in recognizing similar image frames. The 2 examples of 20 query images are shown in Fig. 3.

Each experimental unit was executed by submitting a query image to the program. The best 10 searching results will then be returned. In addition to the experiment, each

Fig. 3. 2 examples of 20 query images. Dog image is 351st frame of a video stream, and fish image is 81st frame of a video stream.

Fig. 4. Examples of query results from video streams for dog image (in Fig. 3) input.

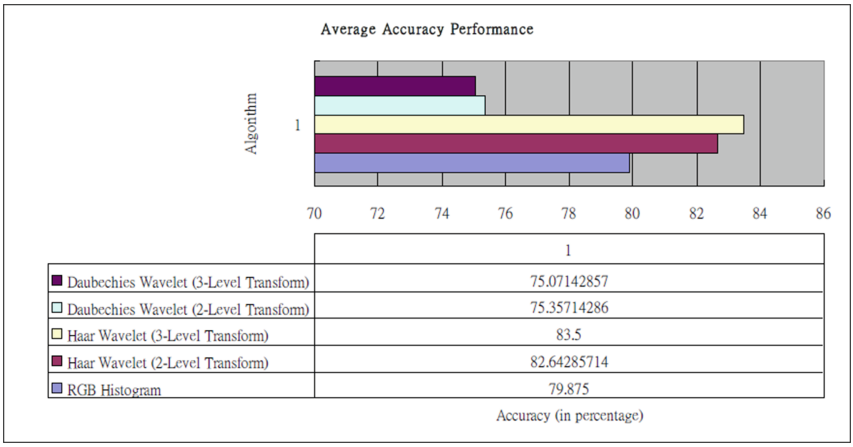


Fig. 5. Average accuracy performance of each algorithm.

experimental result showed in this section specifies the retrieval algorithm used, as well as the clustering technique that was being employed. Figure 4 shows query results of dog image in Fig. 3.

Figure 5 illustrates the average accuracy performance under each of the retrieval algorithm. Both Daubechies wavelet (2-level and 3-level transform) showed less accurate among the 5 algorithms, having only around 75% of accuracy. RGB histogram performed around 4% better than Daubechies wavelet, having around 79% accuracy rate. Haar Wavelet performed the best out of the 5 algorithms, having around 82% to 83% accuracy rate. Haar Wavelet with 3-level transform has the best accuracy rate with approximately 83.5%.

5 Conclusions

Analyzing contents of video streams is essential for implementing ubiquitous video intelligent systems. This paper focuses on an image querying solution from video streams using image features. We have considered RGB histogram and 2 wavelet transform methods like Harr and Daubechies for representing image features.

Experimental results show that the proposed algorithm using 3-level Haar wavelet transform features has the best accuracy rate around 83.5%. We can consider that this level of accuracy is quite useful for developing video analysis applications.

For the future work, optimizations for clustering algorithm and matching/searching algorithm are necessary to improve query performance. Further, development of real-world applications will be required as well.

Acknowledgments. This work was supported by the IT R&D program of MSIP/IITP, Korea [2014-044-020-001, Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis].

References

1. Kang, K., Kwon, Y., Moon, J., Bae, C.: Challenging issues in visual information understanding researches. In: He, X., Luo, S., Tao, D., Xu, C., Yang, J., Hasan, M.A. (eds.) MMM 2015, Part II. LNCS, vol. 8936, pp. 458–469. Springer, Heidelberg (2015)
2. Talukder, K.H., Harada, K.: Haar wavelet based approach for image compression and quality assessment of compressed image. arXiv preprint arXiv:1010.4084 (2010)
3. Daubechies, I.: Orthonormal bases of compactly supported wavelets. Commun. Pure Appl. Math. **41**(7), 909–996 (1988)
4. Wu, Y., Takatsuka, M.: Three dimensional colour pickers. In: Proceedings of the 2005 Asia-Pacific symposium on Information visualisation vol. 45, pp. 107–114. Australian Computer Society, Inc. (2005)
5. Morissette, L., Chartier, S.: The k-means clustering technique: General considerations and implementation in mathematica. Tutorials Quant. Methods Psychol. **9**(1), 15–24 (2013)
6. Memarsadeghi, N., Mount, D.M., Netanyahu, N.S., Le Moigne, J.: A fast implementation of the ISODATA clustering algorithm. Int. J. Comput. Geom. Appl. **17**(01), 71–103 (2007)

Design and Implementation of the Mobile Learning App for Creative Problem Solving Activities

Ji-Hye Bae[1(✉)] and Hyun Lee[2]

[1] Department of History and Culture Contents,
Sun Moon University, Asan-si 31460, Korea
jhbae327@gmail.com
[2] Department of Computer Science and Engineering,
Sun Moon University, Asan-si 31460, Korea
mahyun91@sunmoon.ac.kr

Abstract. To cope with the age of ubiquitous environment, a supply of smart media devices has been indicating a rapid increase due to the development of smart devices as well as the market expansion. The purpose of this study is to propose a method for designing an educational app that uses the mobile learning-based CPS (Creative Problem Solving) model in order to design a mobile learning-based learning environment which contributes to enhancing creative thinking and problem solving ability of the students in college education. The proposed learning support tool is developed in a form of mobile app and is designed to be used as a cognitive tool that enhances the high-dimensional thinking ability of learners through functions such as creative thinking, thinking process, expression method and interactivity in terms of learning activities. As a result of conducting a test by demonstratively applying this app to the actual field of education, the satisfaction rate in participation and learning effect was greater than 90 %.

Keywords: Mobile learning · Creative problem solving · Educational app

1 Introduction

Nowadays, in the educational environment, creativity is considered an important ability to be acquired by individuals involved in the knowledge-based society and it is considered a basic capability to be acquired by individuals living in the modern society to lead innovative developments in the state-of-the-art industry [1]. It is rather important to construct and provide a learning environment that suits the creativity development than to provide an education in which the knowledge is acquired through one-sided memorization or training [2]. Accordingly, a number of researches on educational programs that enhance creativity have been continuously conducted. The CPS (Creative Problem Solving) model is the most representative conceptual model of all approaches and it is related to creativity enhancement. Such CPS model runs a process so that divergent thinking and convergent thinking are developed through a set of stages [3, 4]. CPS model has been diversely used in classes for enhancing creativity,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_121

and the effects have been verified through its application as a learning tool. On the other hand, the information communication technologies such as internet and web have been designed so that any service can be approached through the use of conveniently usable digital devices, and the users are now able to use any service by conveniently selecting and installing apps on their smart devices. Such apps as a form of digital contents have been selected and used by a number of users and such transition has been influential on the current educational environment. Therefore, diverse researches on smart device/app-based education methods have been activated [5]. In the field of education, smart devices have been valuably used due to approachability, portability and convenience and the actual educational apps have been actively developed as well. The purpose of this research is to design and develop a mobile learning-based learning support tool model in accordance to the background described above in order to promote students to conduct creative problem solving activities. Such proposed learning support tool will be produced in a form of mobile app and will be designed as a cognitive tool for enhancing the high dimensional thinking ability of learners through diverse creative thinking process activities and functions such as expression method and interactivity.

2 Theoretical Background

2.1 CPS (Creative Problem Solving) Model

The creative problem solving ability is a thinking process in which divergent thinking and critical thinking dynamically interact and create new products or solutions based on domain-general/specific knowledges and functional grounds [6]. Such CPS model was initially proposed by Osborn and then materialized by Parnes. Such model, as shown in Table 1, consists of the following six stages: mess-finding, data-finding, problem-finding, idea-finding solution-finding and acceptance-finding [3, 4]. Out of the programs for enhancing the creative problem solving ability, most of the models that indicated over 90 % success rate were either the creative problem solving model by Osborn-Parnes or the modified version of the same model [7]. The purpose of this study is to design and propose a learning support tool model based on the 'creative problem solving model' by Osborn-Parnes in order to promote students to conduct creative problem solving activities.

2.2 Mobile Learning-Based Learning Tool

Due to a rapid development of mobile devices that uses the next generation ICT convergence technology, the spread of knowledge and acquirement of information have become very prompt and active. Meanwhile, mobile devices including smart phones have been indicating a rapid progress and the number of users as well as the number of apps being released has been indicating an increase as well. As for the educational environment, the preexisting e-learning system-based learning used to display quite a few limitations in the environment where no desktop existed. Accordingly, a mobile learning system featuring the strengths of portability, mobility

Table 1. Definition per stage of CPS model

Process	Stage	Details
Factors for understanding challenges	1: Mess-finding	Finding experience, role, situation and etc.
	2: Data-finding	Collecting facts such as information, impression and emotion. Exploring situations from diverse perspectives
	3: Problem-finding	Stating diverse problems and sub-problems
Factors for creating ideas	4: Idea-finding	Materializing solutions to stated problems and the possibility of such solutions
Factors for preparing for practices	5: Solution-finding	Creating standards for reviewing and evaluating solutions
	6: Acceptance-finding	Considering agreements/disagreements on such solutions. Confirming potential execution stages

and convenience has been attracting the interest and indicating a gradual increase. In addition, a number of app development tools not only allow the non-computer majors to easily approach app development, but also immediately provide learners with the required learning contents as well as the quick updates and feedbacks on questions and learning progresses. The research specified in [8] is a case where a designed and developed real-time/non-real-time Q&A apps were applied to the field of education, and the results of the class activities were accumulated as database for precisely evaluating the learning activities. In the mobile learning environment, for learning activities such as the creative problem solving activity to be effectively performed, it is important to motivate the learners to participate in such activities and therefore the instructors need to consider such factors into developing the class strategies and using the educational app.

3 Designing of Learning Support Tool for CPS Activities

In this chapter, the designing process involved in the development of learning support tool for promoting creative problem solving activities will be proposed. The overall structure of the system, as shown in Fig. 1, consists of CPS learning tool server, instructor app and student app. In the development environment, since App Inventor is used through the Windows-based web, the configuration is adjusted to the server environment of App Invertor, and the provided components and event handlers are used to develop instructor app and learner app. The instructor app features the provider mode that provides learning activities, and the learner app features the learning activity mode that provides CPS activities.

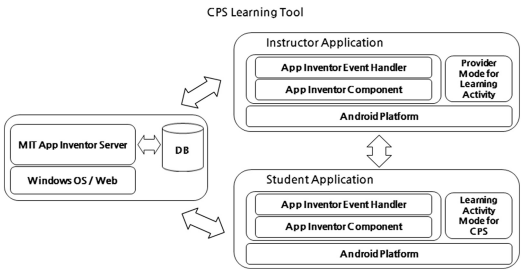


Fig. 1. The overall system structure about CPS learning support tool

4 Implementation of CPS Learning Support Tool

In this chapter, the realization process involved in the development of CPS learning support tool for promoting creative problem solving activities will be proposed based on the designing process described in the earlier chapter. As shown in Fig. 2, this app is designed so that the working screen is divided into the two sections of instructor mode and learner mode.

When the instructor mode is selected, the instructors can login or sign up for a membership, and one room number will be provided per each account. After logging in, the instructors will be able to propose topics for problem solving activities, and will be able to upload related learning data. Such learning data can be uploaded in diverse media formats such as text, image, video and web. In addition, the instructors will be able to intermittently provide feedbacks required by the learners during the activities and will be able to evaluate and score the learning activities performed by the learners as a final stage for completing the instruction activities. In the learner mode, when the room number given to the instructor account is used to login, the app becomes initiated and a group can be created through the team building function. Then the learners will be able to view the problems through the 'view topics for problem solving activities' function. The next activities are creative problem solving activities. The group activities are conducted within the app for solving problems through the following functions: 'challenging problems' for mess-finding, data-finding and problem finding, 'creating ideas' for idea-finding, and 'developing solutions' for solution-finding and acceptance finding.

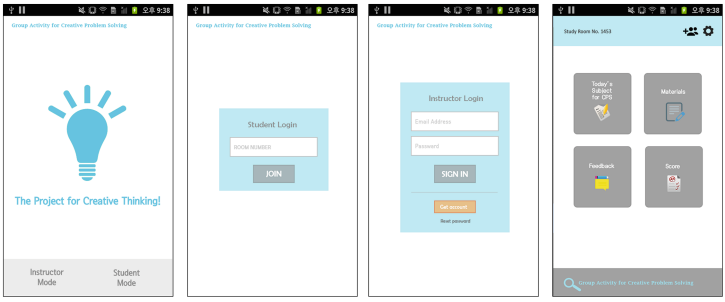


Fig. 2. The sample screens for CPS learning support tool

5 Application to the Field of Education

30 college students participated in the creative problem solving activities demonstrated out by using the CPS learning support tool proposed above. 5 students were put into 1 group and a total of 6 project groups participated in 4 cycles of the online group activities conducted at intervals of 1 week. The project tasks were actively conducted since the participants were familiar with the smart device and the app used in the activities. The topic for the project was set up to be related to the basic knowledge of IT technology. After demonstrating the creative problem solving group activities through the use of CPS learning support tool, a survey was conducted to find out the satisfaction level of the participants. The results are as shown in Table 2.

As a result, more than 90 % of the participants were at least 'Okay' with the use of CPS learning support tool about their participation in group activities, and 90 % of the participants were at least 'Okay' with the learning effects. However, 23 % were 'Unsatisfied' with the convenience of CPS learning support tool and 20 % were

Table 2. Details/Results of survey on use of CPS learning support tool

Q	Details	Response	Reply	
			Respondents	Rate (%)
1	Did the use of CPS learning support tool enhance your participation in group activities?	Very Satisfied/Satisfied	17	57 %
		Okay	10	33 %
		Unsatisfied/Very unsatisfied	3	10 %
2	Did the use of CPS learning support tool contribute to your creative thinking activities?	Very Satisfied/Satisfied	15	50 %
		Okay	11	37 %
		Unsatisfied/Very unsatisfied	4	13 %
3	Was it convenient to use CPS learning support tool?	Very Satisfied/Satisfied	11	37 %
		Okay	12	40 %
		Unsatisfied/Very unsatisfied	7	23 %
4	Were you able to effectively communicate with your colleagues through the use of CPS learning support tool?	Very Satisfied/Satisfied	14	47 %
		Okay	10	33 %
		Unsatisfied/Very unsatisfied	6	20 %
5	Did the use of CPS learning support tool in group activities indicate any learning effect?	Very Satisfied/Satisfied	15	50 %
		Okay	12	40 %
		Unsatisfied/Very unsatisfied	3	10 %

'Unsatisfied' with the communication among group colleagues. 87 % of the participants were at least 'Okay' with the contribution of this CPS learning support tool to their creative thinking activities.

6 Conclusion

In this study, a mobile learning-based learning support tool model is designed and proposed to promote students to conduct creative problem solving activities. The proposed learning support tool is developed in a form of mobile app and is designed to be used as a cognitive tool that enhances the high-dimensional thinking ability of learners through functions such as creative thinking, thinking process, expression method and interactivity in terms of learning activities. The proposed learning support tool is designed in accordance to the android-based app development environment, and the instructor/learner modes are separately constructed to differentiate the involved functions. In addition, the components and event handlers are separately constructed to be compatible with the respective modes. The app proposed in this study is usable in both offline-online classes and is expected to be more valuably used as a learning support tool in classes based on the designing of learning activities for the actual creative problem solving. In addition, most of the creativity-related apps being released consists of programs focused on enhancing brain activities or experiment activities through the graphic process, but since apps related to creative problem solving activities for teaching-learning activities do not exist, the proposed app in this study is expected to be valuably used in class activities. However, viewing it from the perspective of UI design, the user-based screen design is quite insufficient due to the limited component support provided in the development environment, and there are currently limits in constructing a great volume of database since the functions are limited in supplying the data capacity and designing the database. In addition, as a result of conducting a survey, the rate of participants unsatisfied with the convenience of this app and the rate of participants unsatisfied with the communication among group colleagues were considerably high. This signifies that it is necessary to adjust UI in order to apply more significant differentiation since there are other diverse methods that can be used to conduct group activities and communications. In the future research, it is necessary to supplement the functions to expand the user convenience through designing UI suitable for the user environment. In this study, a simple survey was conducted to measure the usability. However, in the future, the proposed app must be applied to the actual field of education in order to additionally review the educational effects of the proposed app on creativity and problem solving ability of the students.

References

1. Mi, K.Y., Soo, K.Y., Sook, N.S., Min, C.S.: A study on the design of a web-based teaching and learning model and learning environments for creative problem solving in mathematics. Res. Inst. Curric. Instruct. **10**(1), 209–234 (2006)
2. Moravesik, M.: Creativity in science education. Sci. Educ. **65**, 221–227 (1981)
3. Osborn, A.F.: Applied Imagination: Principles and Procedures of Creative Thinking, 3rd edn. Charles Scribner's Sons, New York
4. Parnes, S.J.: Guiding creative action. Gifted Child Q. **21**, 460–476 (1977)
5. Rim, H.: Android App. implementation teaching using App. Inventor for elementary school students. J. Korea Multimedia Soc. **16**(12), 1495–1507 (2013)
6. Kim, J., Jeong, H., Kim, H., Kim, H., Lee, W.: Cognitive components definition of creative problem-solving ability in informatics education. J. Korean Assoc. Comput. Educ. **11**(2), 1–12 (2008)
7. Koo, Y.M., Seo, J.H.: Effects of collaborative reflections using SNS on college student' learning motivation, problem solving competency, and academic achievement in Creative Problem Solving activities. Korean J. Educ. Methodol. Stud. **26**(4), 659–685 (2014)
8. Park, C.-J., Kang, J.-H., Kim, M.-J., Yu, Y.-R., Kim, H.-S., Koh, J.-W.: Design and implementation of real-time/non-real-time question-answer apps by using appinventor. In: The 2012 Summer Conference of Korean Association of Computer Education. pp. 63–66 (2012)

IoT-Based Smart Photo Frame

Ji-Hye Bae[2], In-Hwan Kim[1], Yong-Tae Jeon[1], and Hyun Lee[1(✉)]

[1] Department of Computer Science and Engineering,
Sun Moon Univeristy, Asan-si 31460, Korea
{ihkim, mahyun91}@sunmoon.ac.kr, jyt5814@naver.com
[2] Department of History and Culture Contents, Sun Moon University,
Asan-si 31460, Korea
jhbae327@gmail.com

Abstract. In IT Industry, although 'Digital Picture Frame' has advantage of not only bringing the nostalgic memories of classic desk frame but also using smart way to change the picture, it did not become a new trend. It was because there were several problems in digital picture frame such as an expensive price of most importantly, inconvenience of picture transmission system. In this paper, we provide a solution for picture transmission system, and more than that, we add various functions that could use with pictures such as widget and security function. In particular, we use AWS server in the transmission system, and it allows using data in the Wi-Fi environment to transmit pictures whenever and wherever people want it. The proposed system will be the great solution for the discomfort of the current digital picture frame's transmission system which is using USB or SD card. Also, by using this digital picture frame at home or work.

Keywords: Internet of Things (IoT) · Amazon Web Service (AWS) · Wi-Fi · Smart phone · Application · Security · Widgets · Digital picture frame

1 Introduction

With a development of industry, some people are missing and using the products of analog generation instead of developed products because of beauties and special advantages calling as an analog emotion. For example, we heard music by LP(Long Play Record) with turntable at past but nowadays, technology changed turntable with mp3 player such as Smartphone. Yet somewhere people uses turntable because of its beauty and nostalgic memories.

Shape of smart photo frame in this paper looks like existing shape of it and it will attract feeling of nostalgia. Nowadays, we can input pictures to digital frame by USB memory and SD-Card. But this method is convenient. And nowadays, digital frame includes various functions such as music player, touch slide shower, watch, Calendar. But people don't use these functions often because of discomforts.

Therefore, in this paper, we suggest efficient solution to remove discomfort using USB memory and SD-Card and including functions. In addition, we make security function using infrared light sensor and camera. We implement transmission method using AWS (Amazon Web Service) [1] that is an android application. In android application, we send pictures to AWS using Multipart [2] and AWS shows pictures to

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_122

web page organizing of HTML [3]. Digital frame includes small computer called by Raspberry Pi [4] and connects to web page using Chrome [5] and shows pictures using display. Unlike other digital frame, we place widget functions in the corner of one web page. If movement is detected, infrared light sensor catches and takes a picture. We suggest three discriminations such as transmission method, widget function, and security function and overcome the problems of current digital frame.

This paper is organized as follows. In Sect. 2, we introduce the domestic and foreign related research. In Sect. 3, we suggest functional review of this research. In Sect. 4, it is performed experiments and analysis. Finally, in Sect. 5, we conclude this paper.

2 Transmission Methods

In some cases, such as SAMSUNG [6], LG [7], SONY [8] and CEVIA [9], user process using USB memory and SD-Card for transmission method is as follows. First, we save pictures to computer hard disk drive. Second, we copy pictures to USB memory or SD-Card. And finally, we recopy pictures to digital frame. Repeating this step is simpler than analog style of print a picture and insert to frame. But this method is complicated and inconvenient.

In other cases, such as KT [10], SKT [11], Dufri International [12], they use a method of E-mail and MMS. This method is simpler than that of USB memory and SD-Memory. But there are some restrictions. In the E-mail transmission method, we have to memorize E-mail address and turn on internet browser and proceed login process. In the MMS transmission method, we have only on process, but we have to pay a fee.

Thus, we suggest picture transmission method using smartphone application data. This pethood is more cheap and convenient than established method, such as USB memory, SD-Card, E-mail, MMS.

3 Proposed Functional Review of the Study

In this paper, we proposed a smart photo frame that composed of Raspberry-Pi(H/W), AWS(Server), Android application. A core function of this research is composed by three. First function is picture transmission. A method of picture transmission functions is not USB Memory and SD-Card but Wi-Fi transmission system. Second function is Widget. Existing digital picture frame suggests calendar. But when one uses a calendar function, one can't see pictures. The proposed frame shows Widget opaquely on picture and one can see picture and widget at the same time. One can choice and see a widget toggling On/Off switch. Third function is Security. If movement is detected, infrared light sensor detects and takes a picture and send picture to Server. This function can be toggled On/Off too. The combination of CCTV and Picture of captured by security function can prevent crime. This security function is only available to the proposed frame.

3.1 Function of Picture Transmission

Picture transmission process of the frame is as follows. By Multipart form-data form, picture in smartphone is sent to server, AWS. In AWS server, save picture to Server and DB using Multipart form-data of Node.js. To the next, show picture on web page.

This is all that we select picture in gallery in smartphone and touch 'Send' button. So picture transmission process we proposed is simpler than existing process. The proposed frame is cheaper than existing process that need to fee 83 to 200 won per one case.

3.2 Function of Widget

The Widget function process of the frame is figured as shown in Fig. 1. Android smartphone application checks state of 'On/Off' data and saves to DB. After that, HTML verifies DB and expresses Widget of 'On' state. Widget has Time/Date, Weather, Memo, News, and so on. If one toggles Time/Date 'On', display shows time with external Library, 'ClockLink' and date with Date class in JavaScript. For Weather, we use api of National Weather Service in general. But we found bugs in api of National Weather Service and asked to fix. Nevertheless, there is no modification. So we use api of 'Weather UnderGround' and it provides high performance and exact weather information. For Memo, user can save memos using application. At first, system saves memo to DB. Next, system sends specific URL in Server in the form of Json using request module of Node.js. Web page parses data, formed with Json type in specific URL and shows parsed data. For News, widget parses news RSS and shows. In this research, we thought that it is better to show more various kinds of news than specific news. So we applied RSS of 'Google popular news' and shows News as follows. Because of 'Same Origin Policy' [13], we did not RSS Parsing in web but

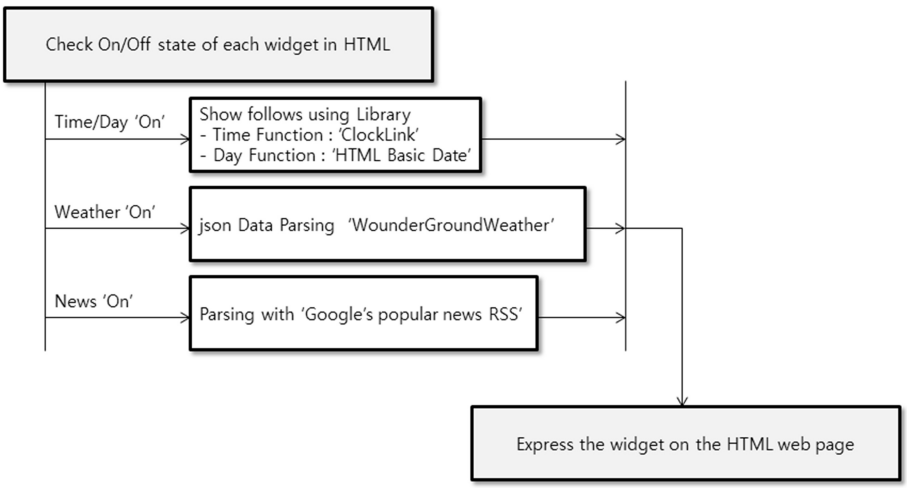


Fig. 1. The process of widget

using cheerio module and request module of Node.js in Server, draws data formed in Json type in specific URL. The system parses data formed in Json type in specific URL using Ajax of Jquery and shows news to user.

3.3 Function of Security

The frame's process of security is figured as shown in Fig. 2. The system saves state of 'On/Off' to DB and Raspberry Pi progresses next step using Python. When a state of security is 'Off', check continuously until 'On' state. When state indicates 'On', system takes a picture using infrared light sensor when senses movement. After that, system sends picture as a Multipart form-data type to Server and returns to first step after 10 s. System receives picture using Multipart form-date of Node.js in Server and saves path & name to DB. At last, android application loads picture in Server by path & name from DB and gives a choice to save picture in the middle of gallery showed by 'Picasso' Library.

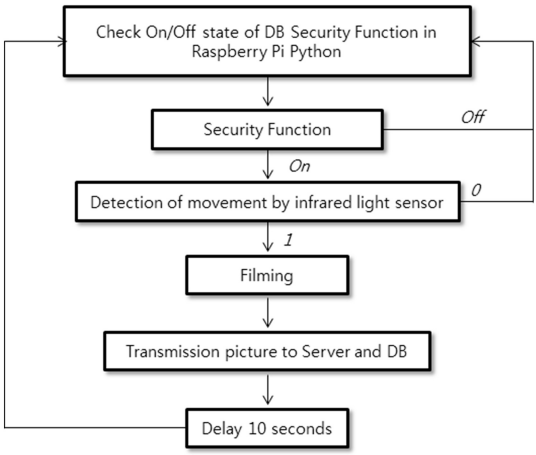


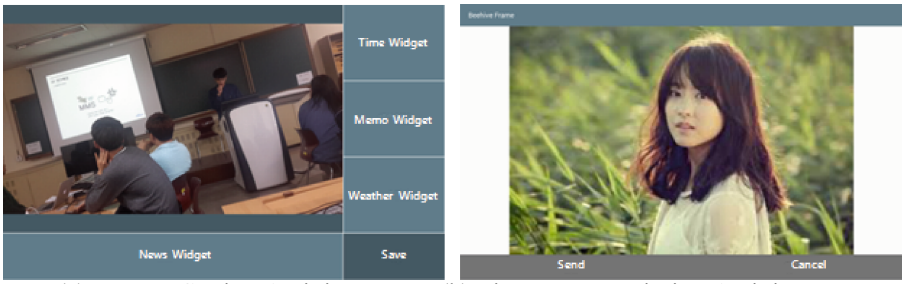
Fig. 2. The process of security

4 Experiment and Analysis

For the experiment of picture transmission, we tested environment of Wi-Fi and LTE-A (Long Term Evolution-Advanced) [14]. After Splash [15] activity is executed, main activity is executed continuously. When you touch 'Frame Setting' area, mode is changed to frame setting activity as shown in Fig. 3(a) and when you push picture area, the gallery of smartphone is activated and if you choice a picture, picture transmission activity will be appear as shown in Fig. 3(b). You can choice between 'Send' and 'Cancel'. If you choice 'Send' button, picture will be sent to Server and picture of your digital frame will be changed.

4.1 Result and Analysis

As shown in Fig. 4, it [16] took 19.515 ms, 9.967 ms, 7.527 ms, 9.358 ms, and 9.453 ms at each time. The transmission took 11.164 ms average. We know that Wi-Fi speed [17] is 11 Mbps and LTE-A speed [18] is 150 Mbps. So we can suppose that it will be faster in LTE-A environment.



(a) Frame Setting Activity (b) Picture Transmission Activity

Fig. 3. An example of experiment

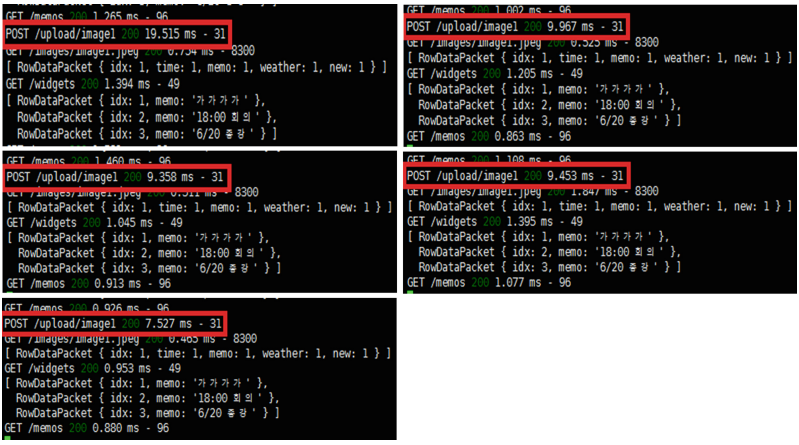


Fig. 4. Speed of picture transmission using Wi-Fi

5 Conclusion

Existing transmission methods such as USB Memory, SD-Card, E-mail, MMS are not efficient because of too much steps and time, inconvenience and cost. Picture transmission process in this paper is efficient because of few steps with application, fast time using Wi-Fi and LTE-A, low cost more than existing method. From now on, direction of research is suggestion of various widgets and applies to various fields.

Acknowledgments. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013R1A1A1075980).

References

1. AWS. https://en.wikipedia.org/wiki/Amazon_Web_Services
2. Multipart. https://en.wikipedia.org/wiki/MIME#Form-Data
3. HTML. https://en.wikipedia.org/wiki/HTML
4. Raspberry-Pi. https://en.wikipedia.org/wiki/Raspberry_Pi
5. Google Chrome. https://en.wikipedia.org/wiki/Google_Chrome
6. SAMSUNG. http://www.zdnet.co.kr/news/news_view.asp?artice_id=00000039171353
7. LG. http://pcpinside.com/1553
8. SONY. http://www.etnews.com/20119140227
9. CEVIA. http://gigglehd.com/zbxe/3190634
10. KT. http://news.joins.com/article/4597206
11. SKT. http://www.hankyung.com/news/app/newsview.php?aid=2005110149118
12. Dufri International. http://m.blog.daum.net/allmart/15651952
13. Same Origin Policy. https://en.wikipedia.org/wiki/Same-origin_policy
14. LTE-A. https://en.wikipedia.org/wiki/LTE_Advanced
15. Splash. https://en.wikipedia.org/wiki/Splash_screen
16. ms. http://terms.naver.com/entry.nhn?docId=1917667&cid=50332&categoryId=50332
17. Wi-Fi speed. http://terms.naver.com/entry.nhn?docId=70666&cid=43667&categoryId=43667
18. LTE-A speed. http://terms.naver.com/entry.nhn?docId=1968165&cid=43667&categoryId=43667

A Fast Algorithm for Generating Virtual Dedicate Network Based on Software-Defined Wide Area Network

Yong-hwan Kim, Buseung Cho, and Dongkyun Kim(✉)

Korea Institute of Science and Technology Information, Daejeon, Korea
{yh.kim086, bscho, mirr}@kisti.re.kr

Abstract. Various needs of new network services have rapidly proliferated over the past few years. To catch up with major requirements, that is, time-to-research and time-to-collaboration, we propose a fast virtual dedicate network generation algorithm based on network abstraction with pruning strategy, unification of multiple links, and an improved spanning tree algorithm.

Keywords: SDN · SD-WAN · Virtual dedicate network · Network abstraction · Exclusive data transmission

1 Introduction

The amount of multimedia & mobile data has been explosively increasing since diverse users, devices, and applications emerged [1]. In addition, innovative intelligent network environment is also being constantly demanded for specific applications such as IoT, big data, cloud computing and so on, based on widely distributed networks for the global repository and access of information [2]. Therefore, new network paradigms have been introduced in terms of the network intelligence, one of which is Software-Defined Wide Area Network (SD-WAN) infrastructure to provide intelligent, flexible, and virtualized network services for geographically distributed things, objects, applications, and users [3], beyond its originating playgrounds in the local area networks [4] such as datacenters, campuses, etc.

However, the conventional user services and applications are very limited because its infrastructure is inherently derived from classical hardware-based, fixed and closed networks, although its users are constantly requiring new network services [5] to guarantee deterministic network performance, user-oriented flexible environment, high availability, security, and virtual network isolations.

In order to achieve those goals, thus, we propose a fast virtual dedicate network (VDN) generation algorithm supporting abovementioned new network demands, for various advanced S&T (Science & Technology) applications and users. To verify and evaluate the proposed algorithm, we also conduct simulation by using Open Network Operating System (ONOS) [6] and Mininet environments. The result shows that the VDN is solely allocated for each S&T user group to meet their major requirements of SD-WAN network, that is, time-to-research and time-to-collaboration.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_123

2 Fast Virtual Dedicate Network Generation Algorithm

In this section, we propose a fast algorithm which generates virtual dedicate network dynamically. The whole system consists of three main building blocks such as physical network, distributed SDN control platform, and VDN application & service. As physical network deployment, the SDN switches are interconnected with multiple links that have differing bandwidth capacities. A VDN can be generated on physical network by allocating network resources such as the switches, ports and links under conditions that (1) VDN participating hosts have to be communicated with each other on the VDN (2) the end-to-end paths between VDN hosts should guarantee required bandwidth. The purpose of VDN application indicates new user interfaces and services along with innovative user experiences, e.g., deterministic network performance and higher security derived from strict virtual network insolation depicted in Fig. 1.

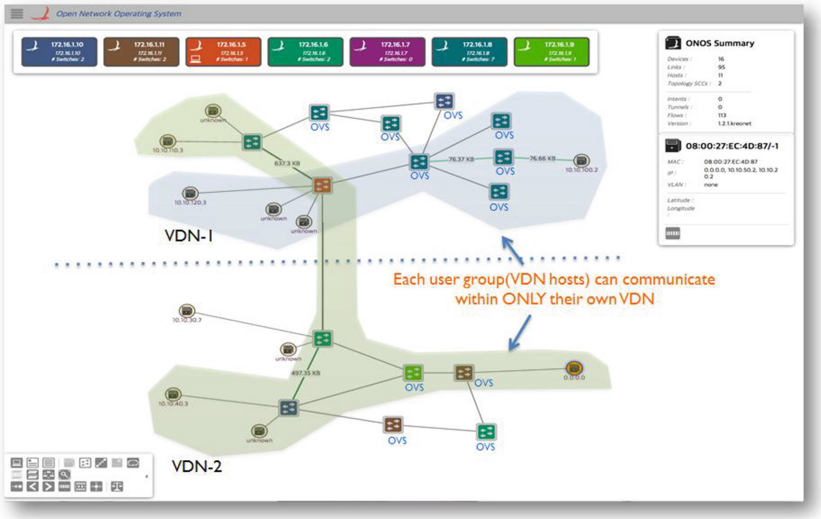


Fig. 1. Examples for isolated VDNs

2.1 Basic Idea of Fast VDN Generation

A VDN can be generated by constructing spanning tree which satisfy above conditions on physical network. The time complexity of constructing tree is usually determined by network scale. In this respect, we try to minimize the VDN generation time as following: (1) network abstraction with pruning strategy, (2) unification of multiple links, and (3) an improved spanning tree algorithm.

First of all, physical network *G(V, E)* is abstracted to subgraph $G'(V', E')$ by pruning unnecessary hosts, switches, and links, where V' is the set of switches having one more available link and E' is the set of available links. *A link is available if* (1) the

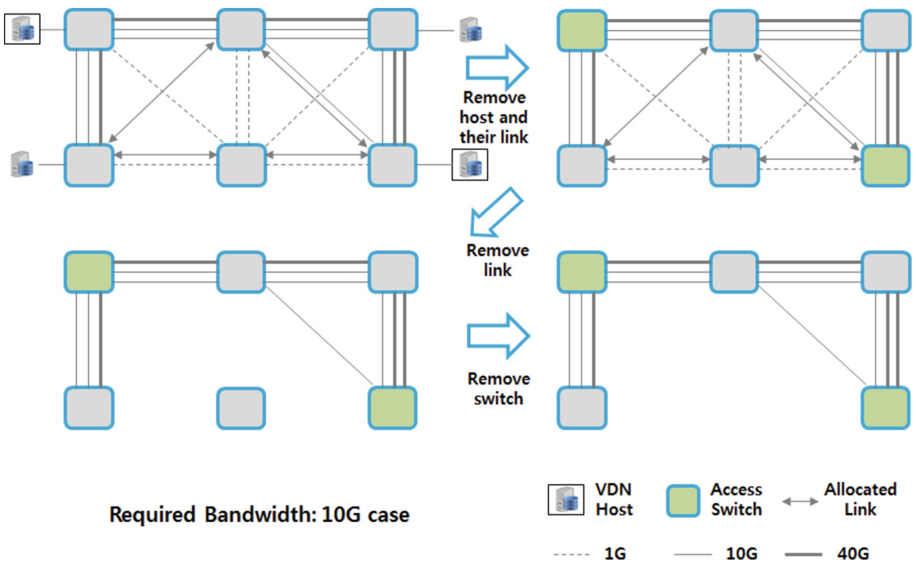


Fig. 2. Network abstraction with pruning strategy

link can support required bandwidth or more (2) and not allocated by other VDNs. Figure 2 shows the example for network abstraction with network pruning in the case of the VDN required 10G bandwidth with two hosts.

The physical network consists of multiple links between switches. However, the only one link among multiple links between switches can be allocated to some VDN. So, we choose one link with minimum bandwidth in $G'(V', E')$ and then remove the other duplicated links before constructing spanning tree for VDN. It is depicted in Fig. 3(a) in succession to Fig. 2.

In order to construct spanning tree T in a general way, the shortest paths for all pair of access switches (denoted as S_{VDN}) related with VDN participating hosts have to be calculated. For constructing T in an efficient way, therefore, we first select a center node of S_{VDN} and then calculate the shortest paths between the center node and S_{VDN}.

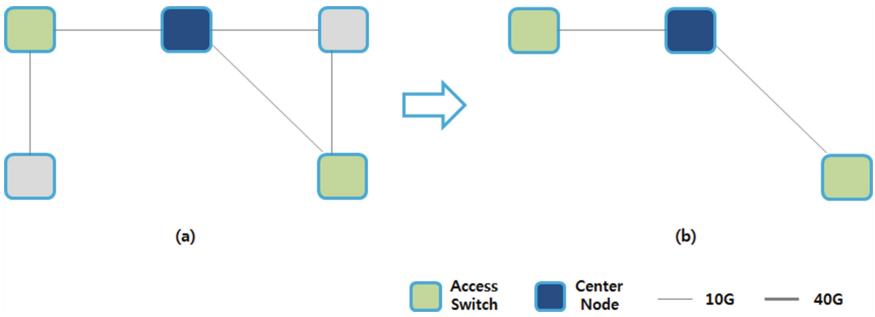


Fig. 3. Unification of multiple links and an improved spanning tree generation approach

This approach can reduce the number of path calculations. The condition of selecting center node c is as following:

$$Min\left(\sum_{v\in V,\notin S_{VDN}}\sum_{s\in S_{VDN}} the\ number\ of\ hop\ count\ (v,\ s)\right) \tag{1}$$

2.2 Proposed Algorithm

In this subsection, we propose a fast VDN generation algorithm by using network abstraction with pruning strategy, unification of multiple links, and an improved spanning tree algorithm. This algorithm returns spanning tree T which is equals to the required VDN. The detailed procedure is shown in Table 1.

Table 1. Fast VDN generation algorithm

Proposed Algorithm ($G(V, E)$, *VDN hosts*, *required BW*)

```
// construct subgraph G'(V', E')
(01) G'(V', E') = G(V, E), S_VDN = ∅;
(02) for each e ∈ E'
(03)   if (BW of e < required BW || e is allocated)
(04)     E' -= e;
(05)   end if
(06) end for
(07) for each v ∈ V'
(08)   if (v ∈ VDN hosts)
(09)     V' -= v, S_VDN = S_VDN ∪ {access switch of v};
(10)     E' -= link between v and access switch of v;
(11)     continue;
(12)   end if
(13)   if (v has no link || type of v is host)
(14)     V' -= v;
(15)   end if
(16) end for
(17) choosing one link with minimum BW between V';
```

```
// construct spanning tree T
(18) min=length of G'(V', E'),sum=0,c= ∅, T= ∅;
(19) for each v ∈ V' and ∉ S_VDN
(20)   for each s ∈ S_VDN
(21)     sum += the number of hop count(v, s);
(22)   end for
(23)   if (sum < min)
(24)     c = v, min = sum;
(25)   end if
(26)   sum = 0;
(27) end for
(28) for each s ∈ S_VDN
(29)   Find a shortest path between c and s;
(30)   T= T∪{v and e on path};
(31) end for
(32) return T
```

2.3 Exclusive Data Transmission on Isolated VDNs

A VDN will be customized for user group to provide innovative user experiences such as deterministic network performance and higher security derived from insolated VDN not whole network. For this, each user group (VDN hosts) can communicate within only their own VDN. That is, the communications between the hosts on different VDN must be restricted. Furthermore, the optimal end-to-end paths between VDN hosts have to be found out by searching only own VDN not whole network. Figure 4 and depicts the examples for exclusive data transmission on isolated VDNs.

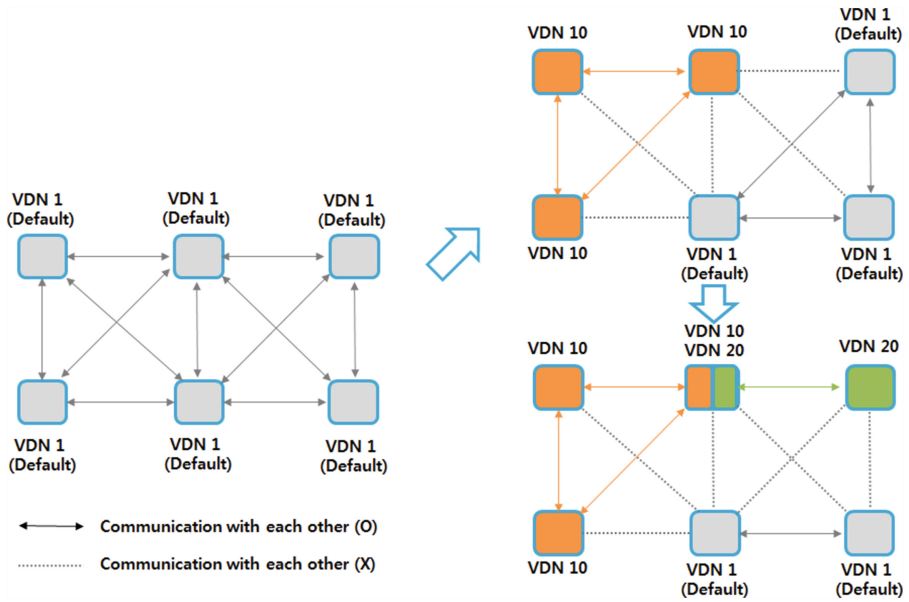


Fig. 4. Examples for exclusive data transmission on isolated VDNs

[Detailed procedure of distributed SDN control platform for data transmission]

- (Step 1) Receive packet in message from a SDN switch
- (Step 2) Parse source (src) MAC address and destination (dst) MAC address in packet
- (Step 3) Find out common VDN including src and dst hosts
- (Step 4-1) Case of existing common VDN: Calculate a shortest path on the VDN and create packet out message including exclusive flow rule according to the path
- (Step 4-2) If not: Create packet out message including filtering rule about the those hosts
- (Step 5) Send packet out message to the corresponding SDN switch

3 Performance Evaluation

In order to verify and evaluate the proposed algorithm, we implement VDN application based on ONOS [6] which is a distributed SDN network operating system. The desired network topology is created using Mininet network emulator.

Based on the above simulation environment, we measure VDN generation time according to the network scale which is determined by the number of switches and links in the network. For each VDN generation, we select the VDN hosts farthest away from each other and the number of VDN hosts is fixed as 4. In case of the network topology, we adopt the hybrid structure of mesh and ring topologies regardless of network scale.

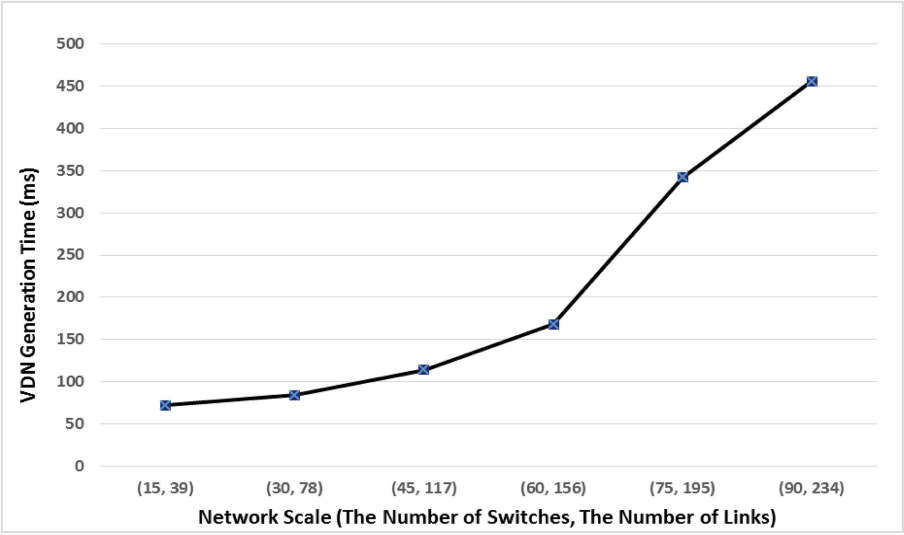


Fig. 5. VDN generation time

Figure 5 indicates the VDN generation time in terms of the network scale, in which from (15, 39) to (90, 234). We can see that the VDN generation time increases linearly as the number of switches and links in the network. In the case of the small network, the completion time of required VDN generation takes less than 100 ms. Even if the network scale is large enough, the completion time do not take more than 500 ms. The result shows that the VDN is solely allocated for each VDN hosts to meet their major requirements of SD-WAN network, that is, time-to-research and time-to-collaboration.

4 Conclusion

In this paper, a fast VDN generation algorithm based on three principal building blocks of SD-WAN is proposed for deterministic network performance and higher security. The performance evaluation shows that the newly developed VDN application has reasonable performance with completion time of service required for time-to-research and time-to-collaboration. We will expect that virtual private & bandwidth-guaranteed network services based on VDN can be good solutions to handle new network paradigms for various advanced S&T applications and users derived from supercomputing, high energy physics, astronomy, bio/genomics, cultural science, etc.

References

1. Cisco Visual Networking Index: Forecast and Methodology. Cisco White Paper, 2014-2019 (2015)
2. Weldon, M.K.: The Future X Network: A Bell Labs Perspective. Taylor & Francis Group, LLC (2016)

3. Manzalini, A., et al.: Software-defined networks for future networks and services. In: White Paper Based on the IEEE Workshop SDN4FNS (2014)
4. Nguyen, V.-G., Kim, Y.-H.: SDN-based enterprise and campus networks: a case of VLAN management. J. Inf. Process. Syst. (2015)
5. Crowell, A., Ng, B.H., Fernandes, E., Prakash, A.: The confinement problem: 40 years later. J. Inf. Process. Syst. (2013)
6. Open Network Operating System. http://onosproject.org/

Detection of Content Changes Based on Deep Neural Networks

Noo-ri Kim[1,4], YunSeok Choi[2,4], HyunSoo Lee[3,4], and Jee-Hyong Lee[1,4(✉)]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Seoul, South Korea
{pd99j, john}@skku.edu
[2] Department of Platform Software, Sungkyunkwan University, Seoul, South Korea
cysuck93@skku.edu
[3] Department of Software, Sungkyunkwan University, Seoul, South Korea
stanlee5@skku.edu
[4] Sungkyunkwan University, 2066 Seobu-ro, Jang an-gu, Suwon-si, Gyeonggi-do 16419, South Korea

Abstract. On R&D projects, automated analysis of implicational and morphological changes between two documents helps managers and researchers to understand projects. However, it is not easy to manually analyze changes between two documents. In this paper, we define text operations which represent changes of texts and make multi-labeled dataset by applying several text operations. Lastly, we propose a method to detect changes of contents. Proposed method represents two documents into an S-matrix first. Next, we use S-matrix as input of Deep Convolutional Neural Networks to identify text operations on the multi-labeled dataset. Experimental results show the effectiveness of our proposed method.

Keywords: Text operation · Multi-label classification · Document modeling · Paragraph vector · Convolutional Neural Network

1 Introduction

On R&D projects, managers and researchers should review products on every milestone to verify progresses. If this process is automated and always performed, it will be helpful to easily understand projects for managers and researchers. In other words, to support content based real-time monitoring on R&D project, an automated analysis is needed. On the automated analysis of text, implicational changes between documents should be a key component for understanding projects. However, it is impossible that managers and researchers check every single document manually because it needs huge amount of costs. There are several software tools which figure out changes between two documents, but they capture morphological changes only. To capture the implicational changes automatically, various text classification methods based on machine learning techniques are proposed [1–4].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_124

Qiu et al. [1] proposed a supervised method for paraphrasing recognition based on similarity and dissimilarity between sentences. Socher et al. [2] introduced a method for paraphrase detection based on recursive Autoencoders. It makes word vectors by recursive Autoencoders and uses word vectors as feature of classifiers. Pang et al. [3] proposed a text matching method based on words' vectors. It makes similarity matrix based on word vectors from two texts and classifies paraphrased text by using similarity matrix. However, this method focuses only on similarities of sentences. Therefore, it is insufficient to identify various changes on texts. Shen et al. [4] proposed a model to capture the relations between questions and answers. It uses S-matrix to contain both lexical and sequential information. It captures the relation between questions and answers which is more difficult than detecting paraphrased sentences. However, this method still considers the relations between two sentences only.

These classification methods have solved various text classification problems. However, since diverse changes can actually occur in a document, there should be at least one or more changes between an original document and a changed document. In other words, to identify changes between documents, we need to consider multi-label problems. In addition, unlike most studies, we need to consider relations between two documents.

In this paper, we define six operations about content changes: *Eliminate*, *Extend*, *Merge*, *Split*, *Rewrite* and *Reorder*. And then, we make multi-labeled dataset and propose a method for detecting content changes using deep neural networks.

The remainder of this paper is organized as follows: Sect. 2 overviews the background. Section 3 describes text operations and a proposed method. Section 4 presents the experimental result and Sect. 5 draws a conclusion.

2 Background

2.1 Paragraph Vector: Distributed Memory

In many natural language processing tasks, there are various attempts to find proper representations that contain text's content. Paragraph Vector: Distributed Memory (PV-DM) [5] is a widely used model for such representation. It predicts the next word based on previous contexts samples from the paragraph. Every paragraph is mapped to a unique vector, and averaged or concatenated with word vector which are also mapped to a unique vector to predict next word in a context. The paragraph vector can be shared across the contexts from the same paragraph, while word vector is shared across all paragraphs. As a result, paragraphs having similar semantics are mapped to similar vectors.

2.2 S-Matrix

S-matrix is used to represent the relations between two vectors. In the natural language process, given two texts, we can figure out the relation between two texts. In general, all words in each text are represented as vectors, and S-matrix is made based on words' vectors. If a text consists of n words and the other text consists of m words, it can be to map two texts into a $n \times m$ matrix, where each element $M_{i,j}$ is calculated by using similarity metric or inner product between first text's i-th word and second text's j-th

word. However, since the number of words varies, matrix cannot be simply fed into machine learning methods. Therefore, on the S-matrix technique, a fixed-large size matrix is used and it is tiled by $n \times m$ matrix repeatedly.

2.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are kinds of deep learning, and CNNs show generally high performance on pattern analysis such as image recognition and image classification. CNNs consist of convolution layers, pooling layers and fully connected layers. Convolution layers can extract different matching patterns. For each layer of CNNs, the k-th kernel such as a mask scans over the input matrix M to generate a feature map. After convolution layers, pooling is applied over the feature maps in convolution layers to capture important features of each feature map. On pooling layers, max pooling is widely used.

3 Proposed Method

Our goal is to detect changes of contents focusing on several text operations. In other words, we try to solve multi-label classification problem. In order to solve this problem, first of all, we define text operations for understanding changes between two documents. And then, we use the S-matrix to regard relations between documents as an image. Lastly, CNNs are adopted, because it shows good performance on image classification. An overview of proposed method is presented in Fig. 1.

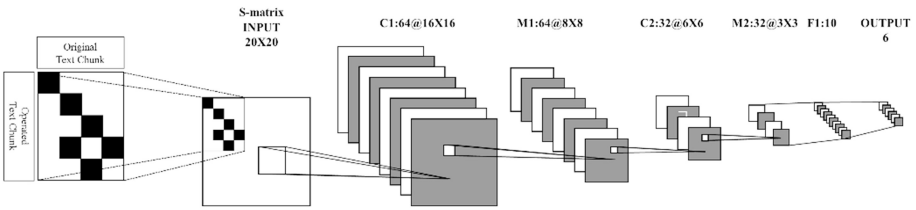


Fig. 1. Overview of detecting method to content changes

3.1 Definitions of Text Operations

In order to understand R&D projects by using changes between documents, we define basic six text operations such as *Eliminate*, *Extend*, *Merge*, *Split*, *Rewrite* and *Reorder*.

Eliminate: *Eliminate* operation represents deletion of one or more sentences. For example, an eliminated text has lesser content than an original text. This operation can lead morphological and implicational changes.

Extend: *Extend* operation represents addition of one or more sentences and it is an opposite of *Eliminate*. If *Extend* operations is applied to a text, changes of morphology and implication will occur.

Merge: *Merge* operation represents concatenation of continuous two sentences as one sentence. If we consider only the number of sentences, *Merge* is similar to *Eliminate*, but it can lead only morphological changes.

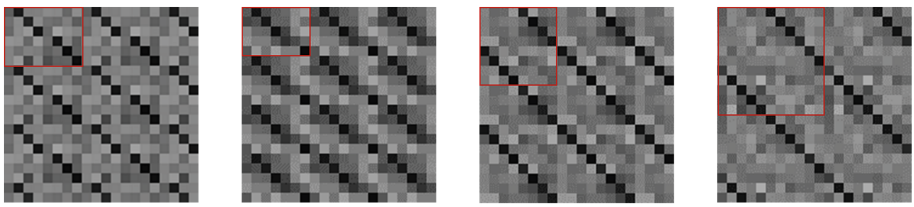
Split: *Split* operation represents division of one sentence as continuous two sentences. It is an opposite of *Merge*. *Split* can lead only a morphological change, not the implicational changes.

Rewrite: *Rewrite* operation represents rewriting or paraphrasing of one or more sentence as the same content. On the *Rewrite* operation, two texts have only morphological changes.

Reorder: *Reorder* operation represents position change of one or more sentences. Therefore, it has only morphological changes.

3.2 Documents Modeling Based on S-Matrix

To model the relations between documents, we apply a documents modeling method based on S-matrix. First of all, we make all sentence vectors from two documents by using PV-DM. And then, S-matrix is filled up by calculating cosine similarities. For example, $M_{1,1}$ in S-matrix is a value of cosine similarity between a first sentence in the original document and a first sentence in the changed document. Figure 2 shows several examples of our dataset.



(a) Eliminate (b) Merge (c) Reorder (d) Rewrite&Reorder

Fig. 2. Example of S-matrixes: *Eliminate*, *Merge*, *Reorder* and *Rewrite* & *Reorder*

In Fig. 2, there are four S-matrixes. Starting point is on the upper left corner and each red box means an actual area between two documents. Horizontal axis means sentences in the original document and vertical axis means sentences in the changed document. Lastly, darker cells mean higher similarities. Figure 2(a) and (b) look similar, because the number of sentences is decreased. However, unlike *Eliminate*, *Merge* has only morphological changes. Therefore, cells around the merged sentences are darker, relatively. Figure 2(c) shows vastly different from previous examples, so it may look easy to distinguish Fig. 2(c). However, as shown in Fig. 2(d), if *Rewrite* and *Reorder* occurs simultaneously, it is still hard to distinguish the changes.

3.3 Detection of Content Changes

In order to detect content changes, we use Convolutional Neural Networks which are a kind of Deep Neural Networks. A structure of CNN consists of convolution layer 1, max pooling layer 1, convolution layer 2, max pooling layer 2, fully connected layer and output layer, sequentially. We use rectified liner unit (ReLU) as activation functions in convolutional layers and sigmoid function as activation function in output layer to solve the multi-label problem.

4 Experiment

4.1 Dataset

We generated a dataset that consists of original documents and changed documents, because there is not suitable dataset to detection of content changes. The dataset was made up using Schutz 2008 dataset [6]. First of all, we separate 1,323 biomedical domain papers into paragraphs. There are 23,658 paragraphs and we regard each paragraph as a document. And then, text operations are applied to every documents, randomly. As a result, there are 23,658 multi-labeled samples. In the dataset, the number of unique label is 41. Label density and label cardinality are 0.3734 and 2.2403 respectively.

4.2 Experiments Design

We experimented two detection models by using Neural Networks as a baseline and the proposed method using Convolutional Neural Networks. On the baseline, it consists of two hidden layers and each layer has 200 nodes with ReLU. We set a vector size of the sentence vector to 300. The size of the S-matrix is set to 20, because it can cover 99.04 % of documents in the dataset. In addition, we apply 5-fold cross validation.

4.3 Results

As an evaluation metric, we used Hamming loss, Accuracy, Precision, Recall, F1-score and Exact. Hamming loss means the number of wrong labels over total number of labels. Table 1 shows results of the baseline and the proposed method. The proposed method shows impressive performance improvement compared to the baseline in all evaluation metric. From this, we found that it is appropriate method to make relation models between two documents as an image and adopt CNNs. We think that these results are sufficient to help managers and researchers for understanding project well by providing changes between documents.

Table 1. Comparison of baseline and proposed method

Metric		Baseline	Proposed method
Hamming loss	(↓)	0.1915	0.1252
Accuracy	(↑)	0.6408	0.7501
Precision	(↑)	0.7850	0.8748
Recall	(↑)	0.7376	0.8224
F1-score	(↑)	0.7606	0.8478
Exact	(↑)	0.3349	0.4759

5 Conclusion

In this paper, we defined six text operations: *Eliminate*, *Extend*, *Merge*, *Split*, *Rewrite* and *Reorder*. Furthermore, we proposed a method for detecting content changes. The proposed method made a similarity matrix by using sentences between two documents, and adopted Deep Convolutional Neural Networks to identify text operations on the multi-labeled dataset. The experimental results demonstrated that the proposed method showed better performance.

In the future, we plan to define more text operations and make model to find complicated relations between documents based on text operations.

Acknowledgements. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030503). And This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (B0101-16-0559)

References

1. Qiu, L., Kan, M.Y., Chua, T.S.: Paraphrase recognition via dissimilarity significance classification. In: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 18–26. Association for Computational Linguistics, Sydney (2006)
2. Socher, R., Huang, E.H., Pennin, J., Manning, C.D., Ng, A.Y.: Dynamic pooling and unfolding recursive autoencoders for paraphrase detection. In: Advances in Neural Information Processing Systems, Granada, pp. 801–809 (2011)
3. Pang, L., Lan, Y., Guo, J., Xu, J., Wan, S., Cheng, X.: Text matching as image recognition. In: 30th AAAI Conference on Artificial Intelligence, Arizona, pp. 2793–2799 (2016)
4. Shen, Y., Rong, W., Sun, Z., Ouyang, Y., Xiong, Z.: Question/answer matching for CQA system via combining lexical and sequential information. In: 29th AAAI Conference on Artificial Intelligence, Austin Texas, pp. 276–281 (2015)
5. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents. In: 31st International Conference on Machine Learning, Beijing, pp. 1188–1196 (2014)
6. Schutz, A.T.: Keyphrase extraction from single documents in the open domain exploiting linguistic and statistical methods. Master's thesis, National University of Ireland (2008)

Resource Allocation in D2D Networks with Location Based Distance Information

Soo Hyeong Kang[1], Pyung Soo Kim[1], Bang Won Seo[2], and Jeong Gon Kim[1(✉)]

[1] Department of Electronic Engineering, Korea Polytechnic University, Siheung Si, Kyunggi Do 429-793, Korea
soo7586@naver.com, {pskim, jgkim}@kpu.ac.kr
[2] Division of Electrical, Electronic and Control Engineering, Kongju National University, Cheonan, Republic of Korea
seobw@kongju.ac.kr

Abstract. Recently, mobile internet traffic has rapidly increased as the huge increase of the smart phone and mobile devices. D2D (Device to Device) is known that it reduce the traffic load of the base station and also improves the reliability of the network performance. However, D2D has a problem that the efficiency decreases as interference is increased. In this paper, we propose a resource allocation scheme to use the resources efficiently when the D2D link share the resources of the cellular network in the uplink. D2D communication utilizes the location information for allocating resources when the eNB know the location of all devices. The proposed scheme not only ensures the performance of the D2D communication but also decrease the computational complexity Simulation results show that the proposed scheme attains the comparable throughput over optimal scheme with a very small computational complexity.

Keywords: D2D communication · Resource management · Location information (GPS) · Computational complexity · Cellular network

1 Introduction

D2D communication is a service for the direct communication between devices without passing through eNB. It is possible to reduce the data traffic and to use limited frequency resources efficiently. Also, it can provide the proximity-based commercial services, such as advertising or marketing, SNS [2].

D2D communication has the advantage of increasing efficiency by reusing resources of cellular networks and to reduce the traffic by spreading the overload of cellular network. However, D2D link and cellular network can cause mutual interference each other between UEs (User Equipment) by using the same resource.

Many researches on D2D communication has been focused on allocating resources in order to prevent the degradation of efficiency that occurs due to the interference. There is a fixed dividing scheme for the available resource for cellular communication and for D2D communication. But, if there are relatively fewer D2D Pairs, it has the possibility to waste resources for full utilization. [3] Besides, there are also schemes

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_125

that ensures performance by reducing the interference on the surrounding through power control. [4, 5] However, those schemes require the higher computational complexity and also consider the more parameter values to be implemented Currently, some techniques to control the interference through a power control has been studied until now while the researches on other methods are relatively fewer than those on power control based schemes.

In this paper, we investigate the resource allocation schemes when the D2D link share the resources of the cellular network in the uplink. The proposed scheme utilizes the location information in order to allocate the resources for D2D communication. Simulation results show that the performance comparison between several allocation schemes for resource allocation. The proposed scheme attains the comparable throughput and the much reduced computational complexity while SINR based scheme achieves the higher throughput and huge increase of computational complexity from the performance comparison.

2 System Model

As shown in Fig. 1, we assume the scenario environment of a single cell to share the same resources on CUE (Cellular User Equipment) and D2D link in uplink for cellular network. We assume that therer are N resources in the cellular network. The number of CUE resources is assumed to be same as the number of CUE and D2D pair use one of the available resources which is already used by CUE.

If typical CUE and D2D link have the same resource, there may exists interference between CUE and D2D link. The symbol, $y_{i,j}$ indicates the interference between CUE i and D2D j. For example, $y_{m,k} = 1$ means that D2D link k share the resource of the CUE m. SINR (Signal-to-Interference-plus-Noise Ratio) of CUE is thus defined as

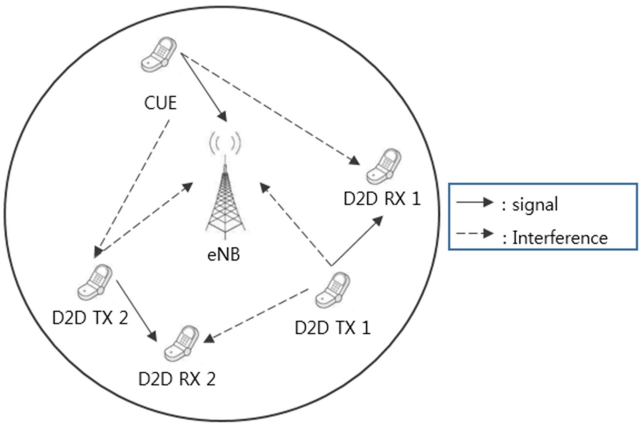


Fig. 1. Interference scenario in D2D based cellular network

$$SINR_{Cellular} = \frac{P_{Cellular} G_{m,eNB}}{P_{noise} + \sum_{k=1}^{d} y_{m,k} P_{D2D} G_{k,eNB}} \quad (1)$$

where $P_{Cellular}$ is the transmission power of the CUE, $G_{m,eNB}$ is the channel gain between CUE m and eNB, P_{noise} is noise power, P_{D2D} is the transmission power of the D2D link and $G_{k,eNB}$ is the channel gain between D2D link k and eNB.

SINR of D2D link is thus defined as

$$SINR_{D2D} = \frac{P_{D2D} G_{D2D\,Pair}}{P_{noise} + \sum_{m=1}^{c} y_{m,k} P_{Cellular} G_{m,k}} \quad (2)$$

where $G_{D2D\ Pair}$ is the channel gain of D2D link between TX and RX and $G_{m,k}$ is the channel gain between CUE m and D2D link k.

The throughput $R_{Cellular}$ and R_{D2D}, for CUE and D2D link are thus defined as

$$R_{Cellular} = log_2(1 + SINR_{Cellular}) \quad (3)$$

$$R_{D2D} = log_2(1 + SINR_{D2D}) \quad (4)$$

where $SINR_{Cellular}$ represent the throughput of the CUE and R_{D2D} represent the throughput of the D2D link, respectively.

3 The Proposed Resource Allocation

3.1 CUE Selection Based on Location Information

When eNB knows the positions of all UEs, all the distances between CUE and D2D link can be calculated from these informations. When we know the locations of each UEs and eNB, the distance between CUE and one of D2D link, $R_{D2D\,RX,CUE}$, can be determined by

$$R_{D2D\,RX,CUE} = \sqrt{(x_{D2D\,RX} - x_{CUE})^2 + (y_{D2D\,RX} - y_{CUE})^2} \quad (5)$$

where the coordinate of CUE and D2D are referred to as (x_{CUE}, y_{CUE}) and $(x_{D2D\,RX}, y_{D2D\,RX})$, respectively. From (5), eNB calculates the values of distance between one D2D link and all the CUEs. Then, eNB sorts the calculated distances between CUE and D2D link in descending order. We select some part of resources from the sorted CUE list since they means the shorter distance between D2D link and CUE uplink and it also indicate that the less interference between them even if they share the resources in the cell.

3.2 D2D Resource Form SINR Comparison

After performing the process in Sect. 3.1, each D2D link can obtain the available lists of CUE resource to be shared. SINR for D2D link is required to consider the interference from CUE as well as the interference from other D2D links which use the same resource with them. The, $SINR_{D2D}$, can be determined by

$$SINR_{D2D} = \frac{P_{D2D}G_{D2D\,Pair}}{P_{noise} + \sum_{m=1}^{c} y_{m,k}(P_{Cellular}G_{m,k} + \sum_{i=1}^{d} y_{m,i}P_{D2D}G_{k,i})} \tag{6}$$

where $G_{k,i}$ is channel gain between D2D link to calculate SINR and other D2D to use same resource.

After calculating SINR of CUE and D2D link for all the possible combination, the maximum throughput of the sum of the throughput of CUE and D2D link can be determined by using (3), (4) and the relative comparion of all the possible values of sum throughput. The maximum throughput is finally shown as

$$Max(\sum_{k=1}^{d} R_k^{D2D} + \sum_{m=1}^{c} R_m^{Cellular}) \tag{7}$$

After calculating the maximum throughput rate, D2D link shares the CUE resource as the throughput rate reached the maximum value in (7).

4 Performance Evaluation and Comparison

In this section, performance evaluation is described for the conventional and proposed scheme through the simulation in a single cell and mixed user traffic environment.

The code for simulation is designed by using C ++. The simulation environment in this paper is presented in Table 1.

Figure 2 shows the total throughput for the four different resource allocation. Random algorithm randomly allocates resource of CUE to D2D link. SINR algorithm utilizes only the SINR when we allocate the resource for the maximum throughput. Location algorithm [6] takes into account only the position information to allocate resource. Finally, proposed scheme use the top 40 % portion of available CUE

Table 1. Simulation environment

Parameter	Value
Cell radius	500 m
Maximum distance between D2D pairs in a pair	100 m
Transmission power of D2D users	23 dBm
Transmission power of Cellular users	46 dBm
P_{noise}	−114 dBm
Number of CUE	5
Pathloss	128.1 + 37.6*log(d/1000)

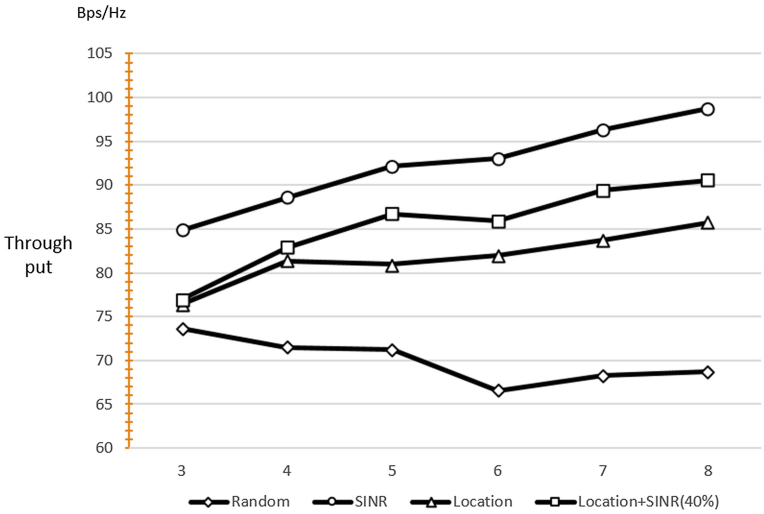


Fig. 2. Total throughput vs the number of D2D links

resources to allocate the resource to D2D pair and applies the SINR scheme to reduce the computational complexity rather than SINR based resource allocation.

The proposed scheme attain the higher throughput over random and location only scheme while it does a little bit lower throughput than the SINR based resource allocation.

The reason for the value of 40 % selction in the proposed scheme is determined by using computer simulation. The throuhghput is increased very sharply when a 40 % of sorted lists are considered for the final selection of optimal resource for D2D pair data transmission.

Table 2 shows the comparison of computational complexity for the SINR and Location based sheme and four different percentage of available resources in the proposed scheme.

The location scheme compute to use only distance information without considering the interference on other D2D link and CUE. SINR based scheme achieves the highest throughput, but the computational complexity will be increased sharply as the number of CUE and D2D pair are increased for the comparison as shown in Table 2.

Table 2. Computational complexity for all the resourece allocation.

Scheme	D2D link = 3	4	5	6	7	8
SINR	125	625	3125	15625	78125	390625
Location	30	40	50	60	70	80
Location + SINR(20 %)	31	41	51	61	71	81
Location + SINR(40 %)	38	56	82	124	198	336
Location + SINR(60 %)	57	121	293	789	2257	6641
Location + SINR(80 %)	94	296	1074	4156	16454	65616

However, the proposed scheme only compares SINR value after selecting 40 % of resources out of the sorted lists based on the distance from D2D pair. Hence, the proposed scheme can reduce the amount of calculation since it has a resource selecting process from all the CUE candidate resources from s in the cell.

Therefore, the proposed scheme can be the best choice for the practical resource allocation in view of the trade off between throughput and computational complexity when we implement this resource allocation in a real cellular networks, which has mixed environment of flexible number of CUE and D2D pairs.

5 Conclusion

In this paper, we investigate the resource allocation in the D2D communication based cellular network. The proposed scheme combines the position information with conventional SINR based resource allocation. It selects some portion of resources based on the distance from CUEs by using the location information and apply the SINR based selection of resources only for these reasonable amount of resource candidates.

Simulation results show that the comparable throughput with SINR based resource allocation and higher throughput than random and location only resource allocation.

Another benefit for the proposed scheme is that it can provide much lower computational complexity over SINR scheme in view of real implementation.

In addition, the resource allocation in this paper assumes only the single cell network and consider the interference from CUE and other D2D pairs. Hence, our future works can be the combination or the additional consideration of multicast, power control, data priority, multi cell and QoS driven D2D data transmission.

Acknowledgment. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2014R1A1A2056387).

References

1. Hong, J.W., Sung, S.I., Park, S.I., Park, C.W., Kim, J.Y., Choi, S.H., Lee, K.B.: D2D communications technology and standardization trends. Mag. IEEE **4**, 77–87 (2013)
2. Corson, M.S., Laroia, R., Li, J., Park, V., Richardson, T., Tsirtsis, G.: Toward proximity aware internetworking. IEEE Wirel. Commun. **17**(6), 26–33 (2010). LNCS
3. Chien, C.-H., Chen, Y.-C., Hsieh, H.-Y: Exploiting spatial reuse gain through joint mode selection and resource allocation for underlay device-to-device communications. In: Wireless Personal Multimedia Communication (WPMC), Taiwan, pp. 24–27 (2012)
4. Zulhasnine, M., Huang, C., Srinivasan, A.: Efficient resource allocation for device-to-device communication underlaying LTE network. In: 2010 IEEE 6th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), pp. 368–375. IEEE (2010)

5. Meshgi, H., Zhao, D., Zheng, R.: Joint channel and power allocation in underlay multicast device-to-device communications. In: 2015 IEEE International Conference on Communications, pp. 2937–2942. IEEE (2015)
6. Rodziewicz, M.: Location-based mode selection and resource allocation in cellular networks with D2D underlay. In: 2015 European Wireless; 21th European Wireless Conference; Proceedings of the VDE, p. 16 (2015)

Design of Corporate Business Card Management System

Seok-heon Ko[1], Gil-mo Yang[2(✉)], and Jun-dong Lee[2]

[1] (Ltd.) AP, 18, Teheran-ro 51-gil, Gangnam-gu, Seoul 05184, Korea
[2] Department of Multimedia Engineering,
Gangneung-Wonju National University,
Wonju-si, Gangwon-do 220-711, Korea
gmyang@gwnu.ac.kr

Abstract. Applying and ordering business card, which are mostly handled manually, cause such problems as work overload, a long time (about a week) till delivery, and hardship in designing several business cards and change designs due to omission in application and the occurrence of miscellaneous work caused by unestablished work process and organizational change which pushes out a large quantity of order at a time. In this respect, the present study suggests a business card management system that is totally computerized so that work process, cost and delivery can be improved. Especially, it connects a company's database with web business card editor to increase to combine convenience and efficiency in applying and ordering business card.

Keywords: Business card request/order · Business card management system · Company DB · Web business card editor

Seok-heon Ko: (Ltd.) AP president.

1 Introduction

Today, printed business card is in common use. In the past, a person who wants to order it had to visit a printing store or an agency, bringing a draft design in person or send a draft design there via facsimile and discuss over a phone. Later, as Internet spread so widely, a system of ordering a business card on the Internet was developed and nowadays, there are a number of printing stores handling business cards on line. In case business cards are ordered on line, many parties in interest are complicatedly involved in multiple processes including an ordering company, a design agency that makes image file, which is time and energy consuming.

As it takes a long time from ordering to delivery due to such complicate procedure of ordering, designing and manufacturing, enterprise business card management system of which process of ordering and manufacturing business card is simple and easy is required.

In this respect, the present study suggests a business card management system that is totally computerized so that work process, cost and delivery can be improved.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_126

Especially, it connects a company's database with web business card editor to increase to combine convenience and efficiency in applying and ordering business card.

2 Literature Review

There are some programs that business cards are designed, printed and ordered on Internet: Business Card MX [1] and Migo 1004 [2]. They provide a variety of templates and design tools so that customers can design business card for themselves and order them. However, their targets are small companies and individual customers so there are limitations for business card management of a large company.

In addition, there are other business card applications [3]: They are based on the method that one enters other's business card ID or telephone number and instantly exchanges business card. In this application, a user gets access to App through QR code and makes electronic business card to exchange or uses NFC function on a smartphone to manage business card. The former is the method that when smartphones contact each other, the owners' business cards are exchanged by NFC [4]. There is also sensor network-based RSSI method. Furthermore, gesture-based electronic business car exchange application is also suggested in an afford to reinforce security. In addition, several studies [5, 6] have been conducted on the methods to manage business card, but few have focused on business card management at a large corporate level.

3 System Design

This section is about ordering and managing system of corporate business card. This system was designed especially to improve work process improvement, reduce cost, and shorten delivery period of business card by combining the database of a company and web business card editor. The business card editor reduces the processes necessary to create business card image by converting business card design to a printable file format. To customize to a company format, this business card management system provides samples in connection with the personnel DB of a company and editor function so that a company can edit and create user information automatically. In addition, once application for business card competes, it creates CMYK file, which is printable format of business card. Therefore, it does not require go through a design agency, but makes a direct order to a printing house.

Figure 1 is the picture of overall system configuration including business card management device. The system consists of user terminal, printing management server and print house.

Business card applicant's terminal (100) is the terminal (130) of a manager in charge of corporate business card management in the related department. The manager's terminal (130) collects and manages business card applications (120) from several departments and makes a final approval. Application for business card is made by a corporate business card application provided in printing management server (200). At this time, the manager compares the contents of applied business cards with a corporate personnel DB to detect error and controls the quantity and period of ordering

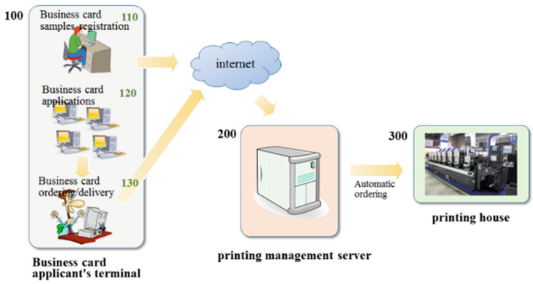


Fig. 1. Corporate business card management system

of business cards. The manager registers the business card samples on printing management (200) in advance or the company's business card beforehand using corporate business card management program provided in printing management server (200).

Figure 2 shows the configuration of printing management server section (200), which is a core part of this system. The server consists of interface section (210), encryption processing section (220), business card ordering management section (230), business card solution section (240), image format conversion section (250), order history analysis section (260), and order processing section (270) and saving section (280).

Interface section (210) provides user interface environment to users ((corporate client (100) and printing house (300) in access to printing management server section (200)) or inputs or outputs data related to business card ordering.

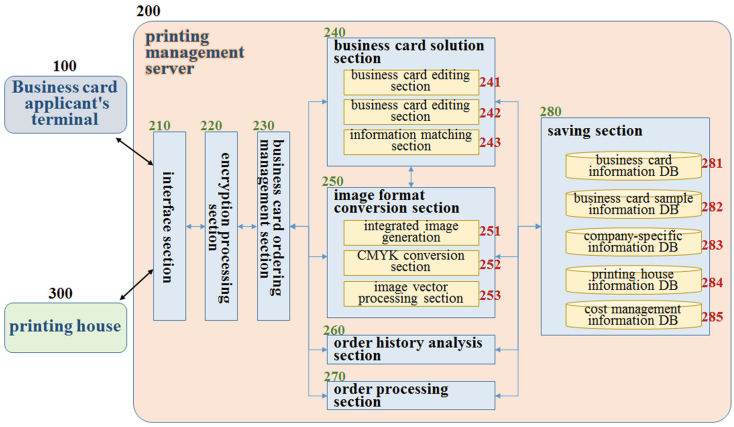


Fig. 2. Detailed view of printing management server

Encryption processing section (220) encodes data output/output data through interface section (210) to protect employees' personal information because the system is connected with a corporate personnel DB, and performs data security by encoding personal information to transfer outside. Business card ordering management section

(230) confirms options that a user selects through interface section (210), searches and retrieves information saved in a storage section (280), and manages ordering, producing and editing of business cards. Business card solution section (240) creates user interface environment, provides production tools to make (design) and edit business card, and creates a draft design combining text and image based on entered information for business card. And business card information of created draft design is compared with personal information records and information of corporate business card samples, which are saved in a storage beforehand, to find error. To enable this, business card solution section (240) includes business card editing section (241) to provide an editing tool for new business card design through interface; business card editing section (242) to provide through interface an editing tool for a user to change his or her personal information based on corporate business card samples registered beforehand by company; and business card information matching section (243) to compare personal information saved in a storage and information of corporate business card samples to detect error. Image format conversion section (250) converts draft business card created through business card solution (240) to CMYK image format and for printing house validation, and creates printable vector image of business card. To enable this, format conversion section (250) consists of integrated image generation (251) to create one integration image through conversion of draft business card to bitmap; CMYK conversion section (252) to create converted CMYK images; and image vector processing section (253) to create printable vector image of business card.

Order history analysis section (260) saves and analyze the status of ordering, production and delivery of business card in connection with corporate client (100) and printing house (300), checks the delivery schedule requested by an applicant, and handles reliability of production of business card.

Order processing section (270) transmits CMYK images, of which format is converted to vector image in format conversion section (250), to printing house (300) when corporate client (100) approves the order of business card and makes a final confirmation of order. Saving section (280) consists of business card information DB (281) by company where business samples registered or designed by company are saved; business card sample information DB (282) where various business samples and their formats (pont, type, form, structure, style, outline, etc.) for the design of business card are saved; company-specific information DB (283) where various information of business card applicants by company (terminal information, contact number, order status, contract information, company information) are saved; printing house information DB (284) where the information of printing house that prints requested business card (address, contact number, quantity of order, credibility, unit cost, etc.) are saved; and business card cost management information DB (285) where information related to cost settlement of ordered and printed business card by company and printing house are saved.

Figure 3 shows the entire process (from order to production) of the proposed business card management system.

Figure 4 is the flowchart that shows the process of creating business card of vector image from a draft business card.

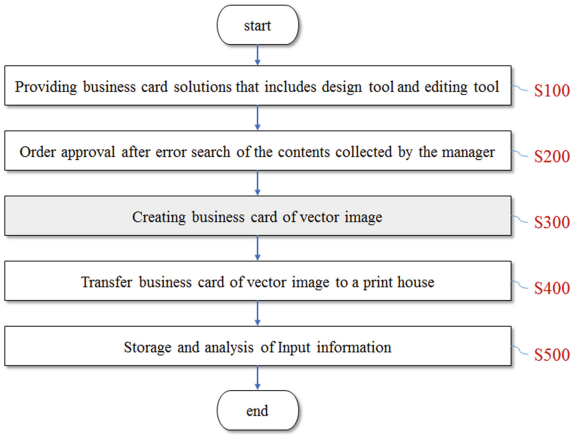


Fig. 3. The flow chart of corporate business card management

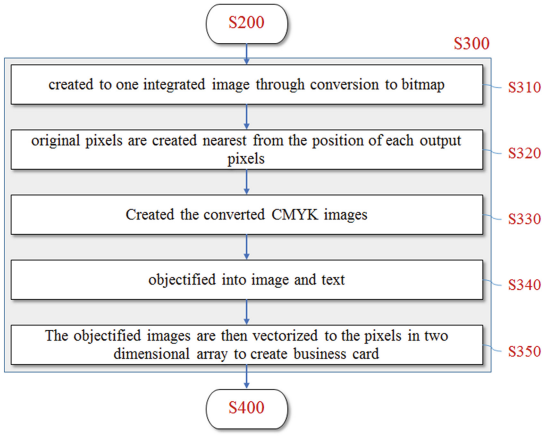


Fig. 4. The flow chart that shows the process of creating business card of vector image from draft business card

First, a draft business card of combined text and image by business card solution section (240) is created to one integrated image through conversion to bitmap by integrated image creation section (251) (S310).

And original pixels are created (S320) nearest from the position of each output pixels, which are detected from the integrated image created through CMYK conversion section (252). These created pixels Modified CMYK is gained through interpolating the values multiplied by weight for the pixels neighboring original pixels (S330).

And then, the converted CMYK images, which are created in image vector processing section (253), are objectified (S340) into image and text by sessionization extraction algorithm. The objectified images are then vectorized to the pixels in two dimensional array to create business card of printable vector images (S350).

The business card of printable vector images created in this way are transmitted to the printing house (300) to make an order (S400). Then, the printing house (300) typesets using the business card of printable vector images ordered from the corporate business card management system (200) and prints out the business cards. The printed business cards are delivered to the client company. In the meantime, printing management server (200) receives the information of complete delivery from the printing house (300) and saves and analyzes ordering, designing, and delivering status of business card, checking the delivery schedule requested by an applicant to secure the reliability of production and delivery of business card (S500).

Please refer to [7] for the building of corporate business card management system.

4 Conclusion

As explained above, the business card management system proposed in this study has effects as follows. First, it can improve work process, reduce cost and shorten delivery schedule by automating applying and ordering of corporate business card through connecting company personnel DB and web business card editor. Second, it can reduce manual works and shorten the reception of business card for printing by converting a draft design to CMYK image format used in a printing house as well as for printing house verification once a corporate completes ordering business cards. Third, it skips the process of format conversion of CMYK images and individual verification by an agency, so it can reduce labor and time to deliver business card to a printing house to almost zero. Fourth, quickened and simplified reception and procedure relieve work load in case a large quantity of order. Fifth, a user (applicant of business card) can select business card format as he or she pleases and makes an easier order of business card.

References

1. http://www.mojosoft-software.com/kr/
2. http://www.migo1004.co.kr
3. http://www.qrook.com
4. Ko, K.-A., Seo, H.-E., Nam, Y.: A NFC-based business card management system for secure many-to-many communication. J. Internet Comput. Serv. (JICS) **16**(3), 13–20 (2015)
5. Kim, M.-R., Park, Y.-H.: A hybrid mobile business card production and management system. J. Korea Multimedia Soc. **11**(1), 117–128 (2008)
6. Shin, H., Kim, C.: A study on performance improvement of business card recognition in mobile environments. J. Korea Inst. Inf. Commun. Eng. **18**(2), 318–328 (2014)
7. PETA information communication. http://www.petait.co.kr

The OpenWRT's Random Number Generator Designed Like /dev/urandom and Its Vulnerability

Dongchang Yoo[1] and Yongjin Yeom[1,2(✉)]

[1] Department of Mathematics, Kookmin University,
77 Jeongneung-Ro, Seongbuk-Gu, Seoul 136-702, Korea
{yoodc8l67, salt}@kookmin.ac.kr
[2] Department of Financial Information Security, Kookmin University,
77 Jeongneung-Ro, Seongbuk-Gu, Seoul 136-702, Korea

Abstract. OpenWRT is an open source router firmware based on embedded Linux that uses a dedicated random number generator for the WPA/WPA2 authentication protocol. Its main purpose is to generate a nonce that can be used in a WPA/WPA2 handshake. If the output of the random number generator can be predicted by an attacker, the relevant protocol will not be able to authenticate securely. According to previous studies on Linux, it is well known that a random number generator implemented in an embedded Linux environment does not collect sufficient entropy from noise sources. The lack of entropy increases the potential vulnerability for the random number generator and the Linux protocol. Therefore, we analyzed the WPA/WPA2 authentication protocol and its random number generator. In our results, we point out some potential cryptographic weaknesses and vulnerabilities of the OpenWRT random number generator.

Keywords: Embedded linux · OpenWRT · Random number generator · Entropy

1 Introduction

In today's world, wireless routers are commonly found in public places, such as Internet cafés and hotel lobbies. Many people join these networks as part of their daily lives. Because wireless routers are inexpensive and can be easily configured without any special knowledge, they are often installed by non-experts. However, an inexpensive device may not contain strong security functions, such as a random number generator. This becomes a major issue when the designer of a cryptographic protocol assumes that the target system is able to generate random numbers with full entropy. In an environment where the target system does not have adequate security functions, there can be significant security breaches, and are most likely caused by weak random numbers.

In this paper, we investigate the structure of the OpenWRT random number generator. Because it is generally implemented in a tiny embedded system, its random number generator lacks the typical noise sources of non-embedded systems. According to a previous research paper [1], the Linux random number generator, both /dev/random and /dev/urandom, may expose some vulnerabilities in an embedded environment like

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_127

the OpenWRT. Embedded systems are inherently vulnerable because of the lack of any human interfaces, such as key-press and mouse movement information. Even worse, embedded systems generally do not contain a spinning disk. These are all used in the Linux kernel random number generator as important noise sources.

We point out an inappropriate design of the random number generator used in the WPA/WPA2 protocol, specifically, entropy collection from noise sources. In addition, we propose a new noise source that can provide additional entropy and be used in a Linux environment.

This paper covers the following topics in the following sections:

- The security analysis of the WPA/WPA2 protocol is described in Sect. 2. In Sect. 2.1, we identify a potential weakness with an attack scenario originated by sending a nonce without encryption.
- The specification and security analysis of the random number generator used in the WPA/WPA2 protocol is covered in Sect. 3.

The preliminary work on this topic is presented in [2]; this paper is a comprehensive and extended version. We delve into more detail of this subject and extract and analyze experimental data harvested from the monitoring function we inserted in OpenWRT for logging the actual noise source of the random number generator.

2 WPA/WPA-PSK Authentication Protocol

When a smart device user requests a connection to the Internet via a wireless router, a mutual authentication is required before the router provides Internet service. The OpenWRT packaged daemon, hostapd, facilitates the cryptographic function including user authentication. In fact, the WPA/WPA2 protocol implemented in hostapd works this way. There are two versions of WPA/WPA2 - PSK, and Enterprise. The WPA/WPA2-Enterprise version includes an additional authentication server, a radius server, and provides a well-designed, strong authentication. Therefore, we will focus on the vulnerability of WPA/WPA2-PSK.

2.1 WPA/WPA-PSK

The authentication protocol WPA/WPA2-PSK [3] implemented in the hostapd enables a client and a router to authenticate each other through a 4-way handshake. Before handshaking, they must pre-share a single password created using English alphabet characters, numbers, or a combination of both. WPA/WPA-PSK performs mutual authentication by having both the router and client calculate a PMK (Pairwise Master Key) using a PSK (Pre-Shared Key). Next, a router and client send a message encrypted with the symmetric key that is shared using a 4-way handshake. The steps of the WPA/WPA2-PSK authentication are described in Fig. 1.

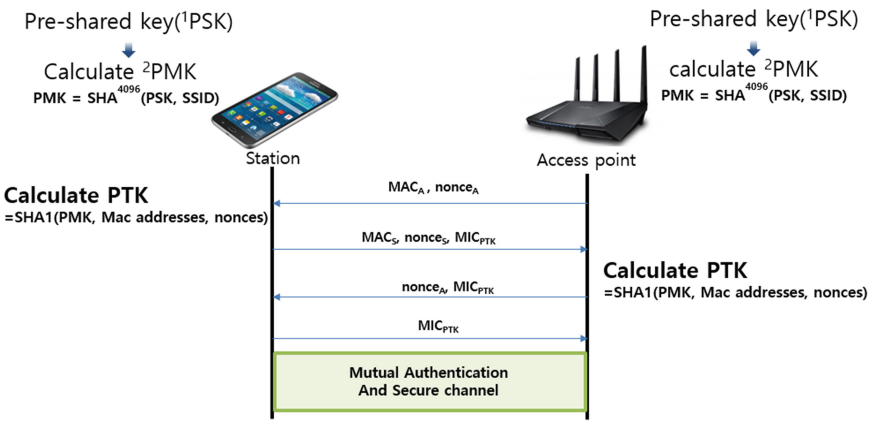


Fig. 1. The WPA/WPA2-PSK protocol with 4-way handshake. In this figure, the left side entity is called a station, or client. The right side entity is the wireless router.

2.2 Cryptographic Analysis of WPA/WPA2-PSK

The WPA/WPA2-PSK protocol has well-known cryptographic vulnerabilities. This protocol was designed under the assumption that all users having the password are honest. In Internet cafés, hotels, and public places, all the customers and visitors can obtain the password easily, including any potential attackers. Therefore, our attack model assumes that an attacker knows the correct password. Furthermore, this attacker can calculate the encryption keys of other users who have connected with the router because all the parameters for deriving encryption keys are public, except for the password. Unfortunately, it is infeasible to overcome this vulnerability. We can only propose that the WPA/WPA2-PSK protocol is no longer used in public places. Otherwise, the router administrator must check face-to-face to determine whether the user is an attacker (Fig. 2).

Fig. 2. An example of poor WPA/WPA2 password management in a public place. The password is posted openly on the table and is too simple.

Another security issue is that a nonce is sent openly in the WPA/WPA2-PSK protocol. In general, a nonce is not dealt with as a secret parameter; it only needs to vary over time. However, if the distribution of a nonce is biased, or is predictable, then the protocol cannot achieve the desired security. The WPA/WPA2-PSK protocol cannot avoid this issue when implemented in an embedded environment.

3 The Random Number Generator of the WPA/WPA2

The Linux environment contains a kernel level Random Number Generator (RNG), also known as /dev/random and /dev/urandom. However, research has shown that the Linux kernel RNG is vulnerable when used in an embedded Linux system [4]. The WPA random number generator was installed and used in the WPA/WPA2 protocol in order to provide more entropy and make up for the potential weakness of the Linux kernel random number generator. We analyzed the source code of hostapd-2016-01-15 in OpenWRT Chaos Calmer 15.05.1, the latest, stable version [5]. Consequently, we found this random number generator implemented in the path ../src/crypto/random.c.

How much more secure is the random number when using an additional RNG?

3.1 Specification of WPA's RNG

The WPA random number generator has three components - noise collection, number generation, and a single entropy pool. All components are similar to the corresponding parts of the Linux kernel random number generator.

Noise sources are mixed into the entropy pool using the mixing function. Notice that the noise source is independent of the Linux kernel random number generator. When a user or program asks to generate random numbers, the extraction function generates a random number using the entropy pool in secret. The following figure presents the structure of the WPA random number generator process (Fig. 3).

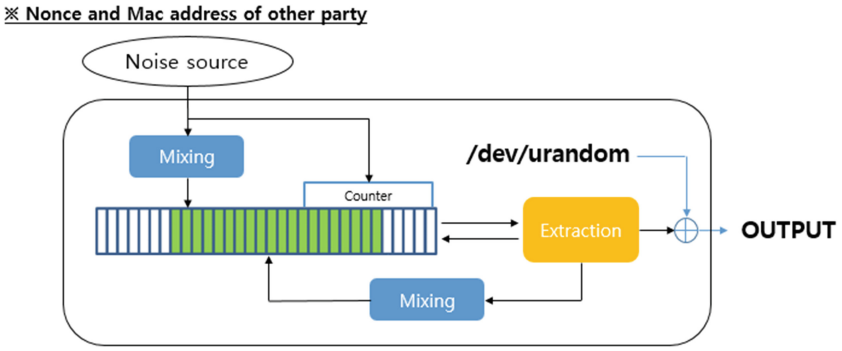


Fig. 3. WPA's random number generator structure. Notice that the random number generator provides the final random numbers based on an output of XOR operation of extraction function and the /dev/urandom device.

3.1.1 Structural Differences Between WPA's RNG and Linux Kernel RNG

The WPA's random number generator was implemented based on the source code of the Linux kernel random number generator [6]. After analyzing the source code, we discovered that the two source codes are almost identical. In this section, we explain the primary differences between the two RNGs.

The WPA's RNG has a single entropy pool while the Linux kernel RNG has 3 entropy pools - 1 input pool and 2 output pools. The output pool is called a blocking pool and is used when creating cryptographic random numbers, as long as the entropy of the input pool is sufficient. Otherwise, output is blocked until enough entropy is accumulated. The other output pool is connected to /dev/urandom and always produces a random number for the user. Therefore, when the entropy of the input pool is insufficient, the random numbers generated are potentially vulnerable. The WPA's random number generator works like /dev/urandom. Despite having little entropy, it creates random numbers without blocking.

A second point is a difference between primitives of each extraction function. A keyed hash function is used in the Linux kernel random number generator while using the SHA-1 cryptographic hash function [7] in the Linux kernel random number generator. We can see that the designers have tried to increase the security of the WPA protocol through the WPA random number generator.

3.2 Security Analysis of the WPA Random Number Generator

We observed the collection process of the WPA random number generator by adding a monitor function to its source code. We then checked the type and quantity of noise sources. The WPA random number generator uses a nonce that is generated by itself, or sent by the other party in the WPA/WPA2-PSK protocol with the opposing MAC address. After collecting data from a noise source, the noise and the time it is collected are mixed into the entropy pool. Because an attacker continuously observing the protocol session can check all entropy sources and the time when the nonce is sent, additional time complexity will increase only slightly. In conclusion, the WPA's random number generator does not increase the security strength of random number generation. That is, using the Linux kernel random number generator with an additional implementation, the WPA's random number generator is not significantly different cryptographically (Fig. 4).

```
   IEEE 802.1X: version=1 type=3 length=121
WPA: Received Key Nonce - hexdump(len=32): 2d 5f 6a ff 57 fa bf d1 00 b9 05 cd e8 cb b9 a4
cd 20 34 80 a4 07 6f e1 7f 08 07 bf 1e fc 08 1a
WPA: Received Replay Counter - hexdump(len=8): 00 00 00 00 00 00 00 01


Im in Adding Function
[Adding]Noise:  - hexdump(len=32): 2d 5f 6a ff 57 fa bf d1 00 b9 05 cd e8 cb b9 a4 cd 20 34
80 a4 07 6f e1 7f 08 07 bf 1e fc 08 1a
[Mixing]Input Data:  - hexdump(len=8): 16 c5 66 57 4d 8b 08 00
[Mixing]Input Data:  - hexdump(len=32): 2d 5f 6a ff 57 fa bf d1 00 b9 05 cd e8 cb b9 a4 cd
20 34 80 a4 07 6f e1 7f 08 07 bf 1e fc 08 1a
Entropy Counter : 1
```

Fig. 4. Partial log files collected show the noise source of the WPA/WPA2 random number generator (*underlined hex values*). After a router receives a nonce from a client, it accumulates the nonce in an entropy pool. This operation is invoked just once per client connection.

4 Conclusion

We analyzed the source code of the WPA random number generator working in the hostapd of OpenWRT to determine its structure. Modifying the program to add noise source monitoring, we could derive real noise sources and determine its collection frequency and quality.

The designer of the WPA/WPA2-PSK protocol tried to enhance the security of the random number generator, but it does not make it more difficult for an attacker to gather random number information than by using the Linux kernel random number generator exclusively. Even though the two environments are different, both were designed like the Linux kernel RNG. This is the reason the random number generator fails to enhance the security. Consequently, we propose that a random number generator be designed with consideration of the environment and recommend that it use more varied noise sources.

Acknowledgments. This research was supported by Next-Generation Information Computing Development Program Through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. NRF-2014M3C4A7030648).

References

1. Gutterman, Z., Benny, P., Tzachy R.: Analysis of the linux random number generator. In: 2006 IEEE Symposium on Security and Privacy (S&P 2006). IEEE (2006)
2. Dongchang Y., Yongjin Y.: Security analysis of random number generator in OpenWRT access point. Korean Inst. Commun. Inf. Sci. (2016)
3. IEEE Std 802.11i-2004. http://standards.ieee.org/getieee802/download/802.11i-2004.pdf
4. Mowery, K., et al.: Welcome to the entropics: boot-time entropy in embedded devices. In: 2013 IEEE Symposium on Security and Privacy (SP). IEEE (2013)
5. OpenWRT Chaos Calmer 15.05.1. released: Mon, 16 March 2016. https://downloads.openwrt.org/
6. Lacharme, P., Rock, A., Strubel, V., Videau. M.: The Linux Pseudorandom Number Generator Revisited. depose sur Cryptology ePrint Archive (http://eprint.iacr.org/) (2012) < hal-01005441>
7. PUB, FIPS: Secure hash standard. Public Law **100,** 235 (1995). Distributed

Implementation of the Block2 Option Transfer for Resource Observing with the CoAPthon Library

Kyoung-Han Kim, Hyun-Kyo Lim, Joo-Seong Heo, and Youn-Hee Han(✉)

Laboratory of Intelligent Networks, Advanced Technology Research Center, Korea University of Technology and Education, Cheonan, Republic of Korea
{goslim, glenn89, chill207, yhhan}@koreatech.ac.kr

Abstract. Utilization of the Constrained Application Protocol (CoAP) that is an important protocol of the Internet of Things (IoT) has increased. So the many associated libraries related with CoAP are also emerging. Among those libraries, CoAPthon, a representative Python-based CoAP library, has advantage in aspect of easy-to-use programming interface to exploit CoAP. However, the current version of CoAPthon has not been implemented correctly in terms of (1) the Block2 option of a block-wise transfer and (2) the transfer for resource observing. In this paper, we implement them and verify the functions in our experiment using the Raspberry PI 2.

Keywords: CoAP · IoT · CoAPthon

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. B0717-16-0033, Study on the multi-dimensional Future Network System Architecture for diversity of Services, Terminals and Networks), and also supported by the MSIP (Ministry of Science, ICT and Future Planning) 2014H1C1A1066391, Korea.

1 Introduction

The Constrained Application Protocol (CoAP), in addition to the communication protocols such as HTTP or MQTT used in the Internet of Things (IoT), has lately attracted considerable attention at development of IoT. CoAP is a protocol standardized by the Internet Engineering Task Force (IETF) and it is a specialized web transfer protocol for use with constrained nodes and constrained (e.g., low-power and lossy) networks [1].

The protocol includes new features such as (1) resource observing [2] to retrieve a representation of a resource and keep this representation updated by the server over a period of time, (2) block-wise transfer [3] to handle the scarcity of memory available on constrained devices for transferring larger payloads.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_128

Currently, there are various CoAP libraries. Among those, Californium [4] is a popular Java CoAP library designed for scalable cloud services, while CoAPthon [6] is a representative Python implementation of the CoAP protocol for easy-to-use programming interface to exploit CoAP. In our research, we have used CoAPthon since the library enables the easy development of IoT application based on the CoAP protocol [6].

However, CoAPthon is incomplete in terms of (1) the Block2 option of a block-wise transfer and (2) the transfer for resource observing. In CoAP, there is a *token* used to match a response with a request. According to the IETF Internet Draft of the block-wise transfer [3], there is no other mentions about the values of tokens for the block-wise transfer, except that the block-wise transfers handle tokens like any other CoAP exchange and a resource requester is free to choose tokens for each exchange of resource information. According to the Internet Draft, therefore, a CoAP server must be able to process a token value regardless of whether the values used for each exchange in a block-wise transfer are identical or not. In the case of a block-wise transfer in CoAPthon, however, a CoAP server just processes the token only when the values of it are identical for all request/response exchanges of a block-wise transfer. So if the resource requester chose a different value for a token delivered by each request messages of a block-wise transfer, a CoAPthon server would not operate properly. In some cases, the size of representation of resources observed by clients may be too large to be transferred in one CoAP message. So the combination of block-wise transfers with resource observing is required. But such a combination is not properly implemented in CoAPthon.

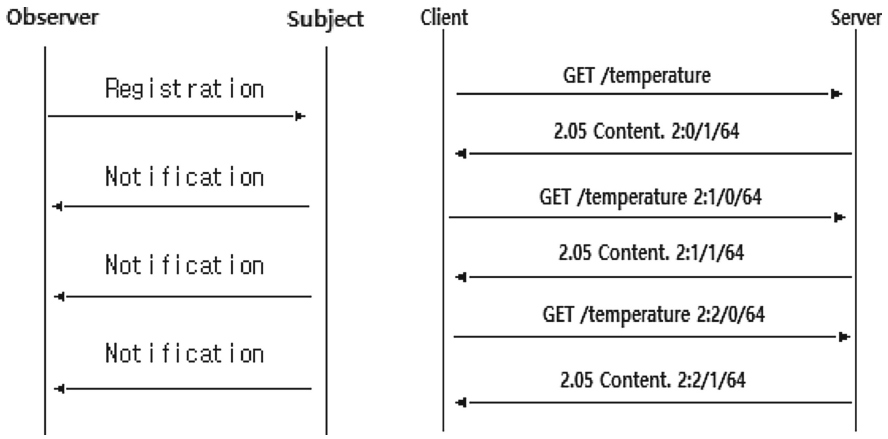
In this paper, we has implemented (1) the Block2 option of a block-wise transfer and (2) the transfer for resource observing within CoAPthon. By using our implementation, we deploy a CoAPthon server on a Raspberry PI 2 equipped with DHT11 temperature and humidity sensor, and verified our implementation.

The remainder of this paper is organized as follows. Section 2 presents related works in terms of CoAPthon, resource observing, and block-wise transfer. Section 3 describes our implementation to make CoAPthon properly in terms of (1) the Block2 option of a block-wise transfer and (2) the transfer for resource observing. Section 4 explains our experiment to verify our implementation. Finally, Sect. 5 concludes this paper.

2 Related Works

A. CoAPthon

The CoAPthon library is built on top of the *Twisted* framework [7]. The *Message Layer* implements the Message sub-layer of the CoAP protocol. It plays an important role to pair messages based on the message header field. The *Request Layer* implements the Request/Response sublayer of the protocol and it handles the CoAP requests and produces its corresponding responses. The *Extension Layer* is the container for all those functionalities that are not included in the basic CoAP specification [6]. Our study focused on the Extension Layer, particularly Observe and Block-wise modules.



(a) The Observer Pattern **(b) Block-wise transfer**

Fig. 1. CoAP extension function

B. Resource Observing

By using the CoAP resource observing, clients can receive notifications from a CoAP server whenever the state of resource, to which the clients registered its interest, comes to change [2]. Figure 1(a) shows a simple example of the resource observing. In the figure, the subject is a resource in the namespace of a CoAP server, an observer is a CoAP client who wants to get a notification from the subject upon the subject's state change. A Registration causes the server to add the client to the list of observers of the resource. A notification is a message that the server sends to each client in the list of observers of the resource whenever the state of a resource changes. A variable-length token value, which can be between 0 and 8 bytes long, is used to match responses to requests.

C. Block-wise Transfer

Basic CoAP messages are exchanged well for the small payloads. But some applications will need to transfer lager payloads. So the protocol provides a minimal way to transfer larger representations in a block-wise fashion. The Block1 Option pertains to the request payload, and the Block2 Option pertains to the response payload. Figure 1 (b) shows a simple example of the block-wise transfer. Block Option fields are composed of three value, NUM, M, and SZX. NUM is Block number, and M is More Flag bit where the payload in the message is the last block if zero sets to it. When one sets to it, it indicates that there are one or more additional blocks available. Finally, SZX indicates the proper size of a block, and its value can be changed by the negotiation between client and server.

3 CoAPthon Extension

The Block-wise transfer in CoAPthon operates based on the *key_token* obtained through the token of request message, client IP and client port. Accordingly, if a block-wise transfer is requested with a unique token value per request to a CoAPthon server, the server would generate an error while responding the first block to every block-wise request since the CoAPthon server expects that the key_token generated by a block-wise request should be identical for all block-wise requests. So, we fixed this problem by generating the same key_token irrespective of the token values delivered from a block-wise request. That is, we implemented a CoAP server so that it uses *uri_path* value of the requested resource instead of the token value when generating key_token.

In the CoAPthon library, on the other hand, there is no implementation needed when a CoAPthon server's resource is observed by a client as well as its representation is transferred through the Block2 option. So, we also fixed the problem while considering the following two cases.

First, let us consider the case that a CoAPthon server needs to transmit a new notification message to its corresponding client due to a periodical delivery policy (or a change of resource state), even if the current notification message composed of block-wise transfer is not yet completely transferred to the client. Figure 2 shows a portion of the *notify* method modified to transmit a CoAPthon server's notification message correctly. As shown in the figure, a new notification message is screened off while the current notification message is being delivered (see the first line in codes shown by the figure). *blockLayer.get_receive* method has function that distinguishes current notification message is being delivered or not.

Second, let us also consider that a CoAPthon server receives a cancel request from a CoAP client for a resource observed by the client, while the server is transmitting the data through the Block2 option for resource observing. In this case, the server must not only remove the client in the subscriber list but also stop operating the block-wise transfer. However, the current CoAPthon library does not support to stop such a block-wise transfer delivered for resource observing, since the server recognizes the cancel request as a block request message and gives answer to that. Then, the client requests the next block so that the block-wise transfer is not stopped.

```
if not(key_token in self._blockLayer.get_receive()):
    transaction.response = None
    transaction = self._requestLayer.receive_request(transaction)
    transaction = self._observeLayer.send_response(transaction )
    transaction = self._blockLayer.send_response(transaction)
    transaction = self._messageLayer.send_response(transaction)
      if transaction.response is not None:
          if transaction.response.type == defines.Types["CON"]:
              self._start_retransmission(transaction, transaction.response)
              self.send_datagram(transaction.response)
```

Fig. 2. The part of notify method

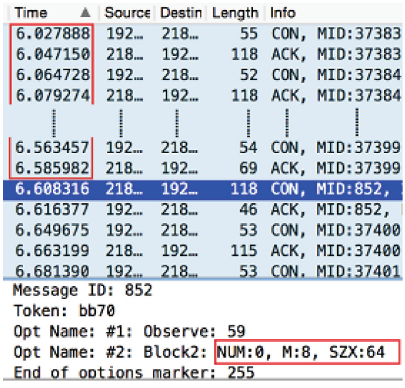
```
if transaction.request.observe == 1:
        m = 0
if key_token in self._block2_observe:
        m = 0
        del self._block2_observe

transaction.response.payload=transaction.response.payload[byte:byte + size]
del transaction.response.block2
transaction.response.block2 = (num, m, size)
self._block2_receive[key_token].byte += size
```

Fig. 3. The part of blocklayer

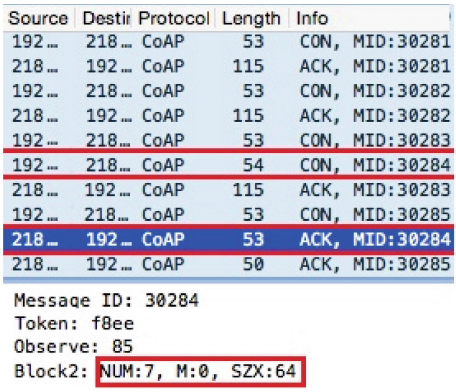
Figure 3 describes how to deal with the problem. As shown in the figure, a CoAPthon server sends the last block with the (N, 0, SZX) option even though a client does not request the last block when the server receives a cancel request while processing a request message.

4 Experiments

To verify our implementation, we built a CoAPthon server on the Raspberry Pi 2 equipped with DHT11, temperature & humidity sensor. As a client, we used The Copper (Cu), Firefox CoAP User-Agent. The user-agent allows users to browse and interact with IoT devices based on the CoAP protocol. The CoAP server sends the current temperature and humidity values obtained through the sensor DHT11 to a registered client. To experiment the Block2 option transfer for resource observing, it is necessary to deliberately increase the size of the data. So we augmented the size of data by including a large number (1024 bytes) of meaningless characters, such as '-', in the data.



(a) New notification message request during the current notification message delivery



(b) Cancel message for the Block2 option transfer

Fig. 4. The Block2 option transfer for resource observing

The Fig. 4(a) shows our experiment about the first case of CoAPthon extension. In the experiment, we configure the periodical (300 ms) delivery policy so that a new message is delivered to the observing client shortly. Then, we make a resource needed about 500 ms to perform a block-wise transfer. As expected, such a new periodic notification message is not delivered until a current block-wise transfer is completed.

Figure 4(b) describes our experiment about the second case of CoAPthon extension. In the figure, a client sends a CoAPthon server a cancel message with MID: 30284 while observing the CoAPthon server's resource through the Block2 option transfer. Then, the CoAPthon server sends a message with the last block option (N, 0, SZX) and MID: 30284 as a response to the cancel message. From the figure, we can know that the client operates well even though sending its cancel message while observing a CoAPthon server's resource through the Block2 option transfer.

5 Conclusions

Among a variety of CoAP library, CoAPthon has advantage to enable the easy development of IoT application based on the CoAP protocol. But the current version of CoAPthon has not been implemented correctly in terms of (1) the Block2 option of a block-wise transfer and (2) the transfer for resource observing, which are needed when a CoAP client wants to receive a large resource representation from a CoAP server whenever the state of resource comes to change. In this paper, we implemented the Block2 option transfer for resource observing correctly about two core cases using CoAPthon. And it was being operated on Raspberry Pi 2 for constrained devices communication. The server communicates with The Copper (Cu) that is Firefox CoAP User-Agent.

References

1. Shelby, Z., Hartke, K., Bormann, C.: The Constrained Application Protocol (CoAP), RFC 7252, June 2014
2. Hartke, K.: Observing Resources in the Constrained Application Protocol (CoAP), RFC7641, September 2015
3. Bormann, C., Shelby, Z.: Block-wise transfers in CoAP, draft-ietf-core-block-20, 29 April 2016
4. Califominum. https://github.com/eclipse/californium.git
5. CoAPthon. https://github.com/Tanganelli/CoApthon
6. Tanganelli, G., Vallati, C., Mingozzi, E.: CoAPthon: easy development of CoAP-based IoT applications with Python. In: IEEE 2nd World Forum on Internet of Things (WF-IoT) (2015)
7. Twisted. https://twistedmatrix.com

Hive-Based Anomaly Detection in Hadoop Log Data Management

Siwoon Son, Myeong-Seon Gil, Seokwoo Yang, and Yang-Sae Moon(✉)

Department of Computer Science, Kangwon National University, Chuncheon, South Korea
{ssw5l76, gils, seokwoo, ysmoon}@kangwon.ac.kr

Abstract. In this paper, we address how to manage and analyze a large volume of log data, which have been difficult to be handled in the traditional computing environment. To handle a large volume of Hadoop log data, which rapidly occur in multiple servers, we present new data storage architecture to efficiently analyze those big log data through Apache Hive. We then design and implement a simple but efficient anomaly detection method, which identifies abnormal status of servers from log data, based on moving average and 3-sigma techniques. We also show effectiveness of the proposed detection method by demonstrating that it properly detects anomalies from Hadoop log data.

Keywords: Anomaly detection · Big data · Log data · Apache hadoop · Apache Hive · Moving average · 3-Sigma

1 Introduction

Recent services and technologies such as social network services, ubiquitous sensor network, and Internet of Things generate a huge volume of data, and big data management and analysis [1, 2] become attractive research fields. Hadoop [3, 4] is a representative software framework to handle such big data efficiently in the distributed environment, and it consists of two major components: (1) Hadoop distributed file system (HDFS) [5] for storing big data in multiple distributed nodes and (2) MapReduce [6] for processing the big data stored in HDFS. However, it is not easy to process big data in Hadoop since we need to convert legacy source codes to complex MapReduce programs. Thus, we generally use Hadoop with other open source solutions, which we call the Hadoop ecosystem, and here we adopt Apache Hive [7] to support the RDBMS-like data management.

In this paper, we address the problem of anomaly detection in Hadoop log data. In general, an actual Hadoop system with multiple nodes generates a huge amount of log data in real time, and it is very important to detect system anomalies by analyzing those log data. In this paper, we present a Hadoop log management system that collects, stores, and analyzes Hadoop log data efficiently. In particular, as an analysis example, we propose a novel anomaly detection method. For this, we first use the Ganglia monitoring system [8] to collect a large volume of Hadoop log data in real or pseudo-real time. We then adopt Apache Hive [7] to store and manage those log data

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_129

efficiently in HDFS. We next propose a simple but efficient anomaly detection method that exploits moving average [9, 10] and 3-sigma [11] techniques. We also actually implement the proposed anomaly detection method by using HiveQL.

We finally empirically validate the effectiveness and efficiency of the proposed system. The experimental result shows that (1) the Hive-based log management system works properly, and (2) the anomaly detection method finds anomalies relatively correctly. These results indicate that the proposed system is a simple but efficient approach to predict the anomalies that might occur in the near future.

2 Related Work

As we mentioned in Sect. 1, Hadoop is composed of two major components: HDFS as a storage model and MapReduce as a computing model. However, implementing MapReduce programs for big data management and analysis in the Hadoop environment is very complicated. Apache Hive [7] has been developed to provide an easy programming model in Hadoop. It presents big data as RDBMS-like table forms on HDFS, and we can access those tables using HiveQL, an SQL-like query language for Hive. More specifically, to process HiveQL queries, Hive first converts those queries into MapReduce algorithms and then executes the algorithms in the Hadoop MapReduce framework.

In general, the size of record items in log data is small, but their generation speed is very fast, and we regard the log data as a typical example of big data. We can collect such log data through a log monitoring or log collecting system, and Ganglia [8] is one of such representative log monitoring systems. In this paper, we use the Ganglia monitoring system to collect Hadoop log data, store the collected log data into HDFS through Hive, and detect Hadoop system anomalies from the stored data. As the technical background of anomaly detection, we exploit moving average [9, 10] and 3-sigma [11] techniques. First, the moving average technique converts a time-series into an average sequence of the fixed time interval, and it is widely used to understand the changing trend of time-series data. Second, the 3-sigma technique is an experiential rule that, if the measured values follow the normal distribution of mean μ and standard deviation σ, 99.7 % of the data values are included in $[\mu-3\sigma, \mu+3\sigma]$. Using these moving average and 3-sigma techniques, we identify the values which are not in $[\mu-3\sigma, \mu+3\sigma]$ as anomalies, where we get μ and σ from the moving averaged sequences.

3 Hive-Based Anomaly Detection from Hadoop Log Data

Figure 1 depicts the proposed Hive-based log management and anomaly detection system. As shown in the figure, the whole system consists of four major components: Hadoop name or data node, Ganglia, Hive, and anomaly detection program, which work as follows. First, the Ganglia monitoring system collects system log data from each Hadoop name or data node and deliver the log data to Hive. Second, the Hive system stores the Hadoop log data into the predefined tables of HDFS. Third, the

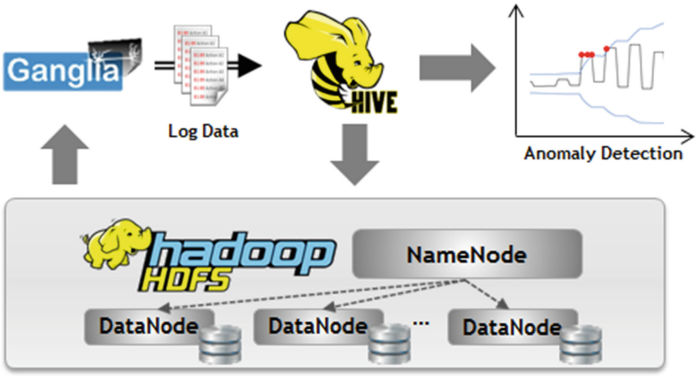


Fig. 1. The proposed Hive-based log data management and anomaly detection system.

anomaly detection program detects system anomalies from the log data stored in tables of Hive and reports the anomalies to users.

Figure 2 shows a part of real log data collected by the Ganglia monitoring system. Ganglia first groups multiple nodes in a cluster to be monitored, and it then collects Hadoop log data from the cluster. Explanation of each column in Fig. 2 is as follows. The first column '`etc`' of each row represents the cluster in which the log data are collected. The second column '`agent1`' represents the hostname of the corresponding server. The third column represents the type of log data, and in this case, the '`bytes_out`' type means the size of network output data. The fourth column represents the measured value. Finally, the fifth column means the measured time. If we rearrange the values of each column into a sequence, we can regard it as a time-series of the specific column.

```
etc,agent1,bytes_out,1403.47,20150114000001
etc,agent1,bytes_out,1403.47,20150114000101
etc,agent1,bytes_out,1126.54,20150114000201
etc,agent1,bytes_out,1126.54,20150114000301
etc,agent1,bytes_out,1126.54,20150114000401
etc,agent1,bytes_out,1126.54,20150114000501
etc,agent1,bytes_out,1126.54,20150114000601
```

Fig. 2. Hadoop log data collected by the Ganglia monitoring system.

As shown in Fig. 2, the log data collected by Ganglia are text files of CSV(comma separated value) format, and we need to convert those text format into table format to be handled in Hive. For this, we first design a table scheme which stores the log data in Hive. Figure 3 shows the Hive table scheme that stores Hadoop log data. As shown in Fig. 3, the table structure follows the CSV formatted log data of Fig. 2, and it consists of five attributes each of which corresponds to each column of Fig. 2.

Fig. 3. A Hive table scheme for storing Hadoop log data.

We now present a HiveQL-based query algorithm that detects system anomalies from the Hadoop log data stored in Hive. Figure 4 shows a query that implements the anomaly detection exploiting moving average and 3-sigma techniques. First, an inner query statement of Lines 6 to 10 computes the moving average and standard deviation from '`log_table`' which stores the actual log data. Here, `MOVING_AVG()` and `MOVING_STD()`, user defined functions (UDFs), compute the moving average and standard deviation, respectively, over a specific window, and the window size can be changed by a given parameter. For simplicity, in this paper we use 1 h, i.e., 60 min, as the window size. Next, an outer query statement of Lines 1 to 4 executes the anomaly detection process by using the computed moving average and standard deviation. More precisely, in Line 2, if the log value is `NULL`, we identify it as an anomaly and return `TRUE`; in Line 4, we return `FALSE` for the other normal values. Line 3 is a core part that determines an anomaly based on the moving average and standard deviation. Here we note that, if a log value is not in [moving average $\pm$ 3$\times$standard deviation], we determine it as an anomaly and return `TRUE`.

```
1) SELECT tt.time_stamp, tt.metric_value,
2) CASE WHEN tt.metric_value IS NULL THEN 'TRUE'
3)      WHEN tt.metric_value NOT BETWEEN tt.moving_avg - 3*tt.moving_std
             AND tt.moving_avg + 3*tt.moving_std THEN 'TRUE'
4)      ELSE 'FALSE' END is_anomal
5) FROM (
6)      SELECT t.time_stamp, t.metric_value,
7)      ROUND(MOVING_AVG(t.metric_value, 60), 3) as moving_avg,
8)      ROUND(MOVING_STD(t.metric_value, 60), 3) as moving_std
9)      FROM log_table as t
10)     ORDER BY time_stamp) tt;
```

Fig. 4. An anomaly detection algorithm exploiting moving average and 3-sigma techniques.

4 Experimental Evaluation

In this section, we evaluate the effectiveness of the proposed anomaly detection method through an actual experiment. We perform the experiment in a Hadoop cluster which is configured by a name node and four data nodes. As software platforms, we use Hadoop-2.6.0 and Hive-1.1.0. We use a '`bytes_out`' file, which contains one day log data collected from the '`agent1`' server in the '`etc`' cluster.

Figure 5 shows a visualization graph of anomaly detection results. In the figure, *x*-axis represents the measured time from 00:00 to 24:00, and *y*-axis represents the

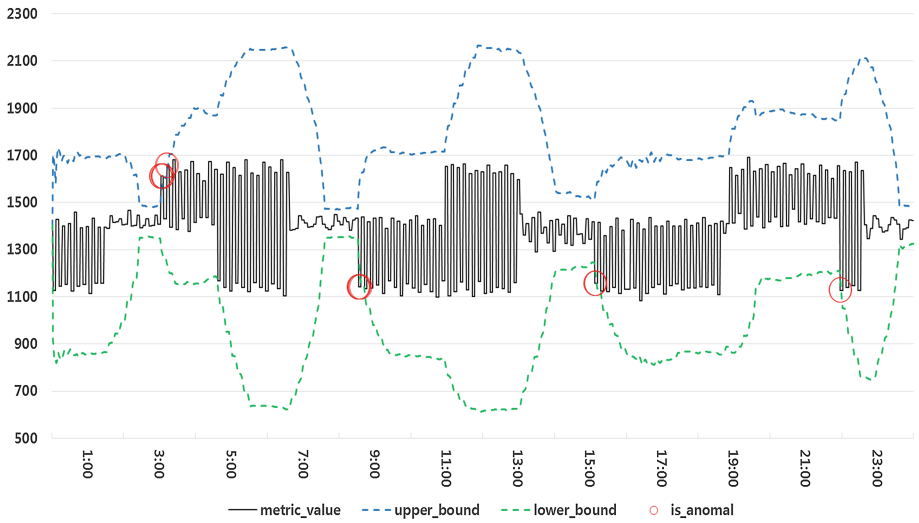


Fig. 5. A visualization chart of anomaly detection results.

measured value, i.e., '`METRIC_VALUE`' in Fig. 3. The solid line represents log values, and two dotted lines represent the 3-sigma band. That is, if the solid line deviates the dotted band, we determine those points are highly to be abnormal, i.e., they might be anomalies. As shown in the figure, we find four abnormal intervals, which are depicted as red circles in the chart. In particular, we note that there are rapid changes in between 07:00 and 09:00, and we can say that there might be some abnormal status at that time.

We now analyze the experimental result of Fig. 5 in more detail. First, our anomaly detection method correctly identifies abnormal points in which the log data values are rapidly changed. For example, the change before 15:00 is not large, but the change around 15:00 is very large, i.e., the log values rapidly increase around 15:00, and we identify a rapid change time point as an anomaly. Next, let us discuss another time interval in between 11:00 and 13:00. There is also a big change at that interval, but the change ratio itself is not large, and thus, we regard those time points as normal changes. However, our anomaly detection method has a problem of reporting unnecessary anomalies in the case where there are rapid changes. For example, seven anomalies are reported around 03:00, and five anomalies are reported around 09:00. These multiple reports might confuse the analyst, and we leave the removal of those multiple reports as the future work.

5 Conclusions

In this paper, we addressed an anomaly detecting solution in the Hadoop environment. For this, we first presented a Hive-based log data management and analysis architecture. We then proposed an anomaly detection method using HiveQL, where we adopted moving average and 3-sigma techniques. We also performed the experiment on the

actual Hadoop log data and visualized the anomaly detection results as a graphical chart. Experimental results showed that our method correctly identified abnormal points as anomalies. As the future work, we will apply the presented method to log data of multiple Hadoop nodes, and redesign the Hive storage structure to improve its maintenance and analysis performance.

Acknowledgement. This work was supported by Institute for Information & communications Technology Promotion(IITP) grant funded by the Korea government(MSIP) (No. R7117-16-0214, Development of an Intelligent Sampling and Filtering Techniques for Purifying Data Streams)

References

1. Rabl, T., Sadoghi, M., Jacobsen, H.-A., Gómez-Villamor, S., Muntés-Mulero, V., Mankowskii, S.: Solving big data challenges for enterprise application performance management. Proc. VLDB Endowment **5**(12), 1724–1735 (2012)
2. Saecker, M., Markl, V.: Big data analytics on modern hardware architectures: a technology survey. In: Aufaure, M.-A., Zimányi, E. (eds.) eBISS 2012. LNBIP, vol. 138, pp. 125–149. Springer, Heidelberg (2013)
3. Hadoop. http://hadoop.apache.org/
4. Lam, C., Warren, J.: Hadoop in Action. Manning Publications, Greenwich (2010)
5. Shvachko, K., Kuang, H., Radia, S., Chansler, R.: The hadoop distributed file system. In: Proceedings of the 26th IEEE Symposium on Mass Storage Systems and Technologies (MSST), Lake Tahoe, Nevada, pp. 1–10, May 2010
6. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
7. Thusoo, A., Sarma, J.S., Jain, N., Shao, Z., Chakka, P., Authony, S., Liu, H., Wyckoff, P., Murthy, R.: Hive: a warehousing solution over a map-reduce framework. Proc. VLDB Endowment **2**(2), 1626–1629 (2009)
8. Ganglia Monitoring System. http://ganglia.info/
9. Lucas, J.M., Saccucci, M.S.: Exponentially weighted moving average control schemes: properties and enhancements. Technometircs **32**(1), 1–29 (1990)
10. Moon, Y.-S., Kim, J.: Efficient moving average transform-based subsequence matching algorithms in time-series databases. Inf. Sci. **177**(23), 5415–5431 (2007)
11. Pukelsheim, F.: The three sigma rule. Am. Stat. **48**(2), 88–91 (1994)

HIM-PRS: A Patent Recommendation System Based on Hierarchical Index-Based MapReduce Framework

Xuhua Rui and Dugki Min(✉)

Department of Computer, Information and Communication Engineering,
Konkuk University, Seoul, South Korea
{abealasd, dkmin}@konkuk.ac.kr

Abstract. Intellectual Property (IP) data, such as patent documents, grows inconceivably in recent years. Therefore, discovering valuable information from those huge number of data becomes a challenge. This paper introduces a novel patent recommendation system called HIM-PRS which is built on top of hierarchical index based big data processing platform. HIM-PRS integrates with linked data to provide an efficient patent recommendation service. Our result shows that HIM-PRS is able to find more semantically similar patents than other systems. Additionally, HIM-PRS launches query jobs at least 2 times faster than original Hadoop MapReduce framework.

Keywords: Patent recommendation service · HIM · BigData Processing · Linked data

1 Introduction

With rapid development of technologies, Intellectual Property (IP) plays a more and more important role in our life. Because it represents the legal power for people to protect their invisible treasure. Last year, more than 7500 patent disputes happened in the U.S., and more than half of them are related with high technology [1]. However, an intellectual property dispute will cost a various of long term lawsuits. For example, Apple Inc. started to sue Samsung Electronics Co., Ltd. in the year of 2010. Finally, various of lawsuits in more than 9 countries are involved in their IP war during the past 5 years [2].

There are a lot of reasons making those disputes so complicated, durable and difficult. The most important reason is the rapidly increased number of patents cannot be analyzed within a short time. In the year of 2000, only 175979 patents have been granted by United States Patent and Trademark Office (USPTO). 4 years ago, the number became 276788. During the past decade, more than 3 million patents has been officially granted by USPTO [3].

Discovering meanings from those amount of data challenges more and more IT companies. USPTO has opened all U.S. patent documents to public and provided a lot of methods to access their database [4]. However, the query performance limits users. From 2010, Google, the global internet information searching service provider,

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_130

co-worked the USPTO to provide their bulk data for public access [5]. Google also opened a patent searching service [6] for helping users to discovery IP information easily.

Since such amount of patent document data has been opened, researchers apply various of approaches to discovery treasures from them. Statistical patent analysis approach [7] has been widely used for simply measuring and visualizing patents based on technology catalog. Citation based patent analysis [8] approach is inspired by information retrieval technology. It is able to present inner-relations of patents in addition. Patent cluster analysis is another patent analysis approach. By utilizing cluster analysis approach, patent documents can be organized as a theme map which describes patents relationship simpler [9]. Furthermore, researchers are interested in utilizing ontology technology to discovery knowledge from patent data [10].

However, there is still not a sufficient solution for analyzing patent big data based on the semantic meaning of patents. In this paper, we will introduce our proposed system, HIM-PRS, which is designed for analyzing patent bigdata semantically.

The rest of the paper is organized as follows: We first show the architecture of HIM-PRS, and afterward, we describe the system working mechanism in Sect. 2. In Sect. 3 we present the experimental results of our proposed system with analysis. Finally, we give the conclusion.

2 HIM-PRS Architecture

In this section, we mainly focus on introducing the design of our proposed patent recommendation system: HIM-PRS.

2.1 HIM: Hierarchical Index-Based MapReduce Framework

The fundamental of our proposed system is a bigdata storage and processing system which is called HIM. HIM is inspired by Google's MapReduce paper [11] and built on top of the most widely used open source bigdata solution Hadoop [12].

The original Hadoop keeps data uploaded by users in separated blocks. However, as time flies, people realized that the plain data storage structure is no longer sufficient for big document analysis. Database technologies performance a good result for data retrieval by using index. However, as the data grows, the index generation time and size become unacceptably. Therefore, many researchers try to integrate the advantages of both technologies, such as HadoopDB [13], HBase [14], Hive [15], Hadoop ++ [16] and BigTable [17]. Inspired by previous researchers' work, we proposed a hierarchical index-based patent data storing mechanism [18].

Based on the previous proposed mechanism, we built our hierarchical index-based MapReduce framework for storing and processing patent documents.

Figure 1 shows the main concept of HIM. Since HIM is built on top of Hadoop, there are 2 kind of servers are involved. A server works as a master node which maintains the entire distributed system(DS) information. A master node is in charge of creating, updating, deleting, backup and recovering the DFS Hyper Index. DFS Hyper

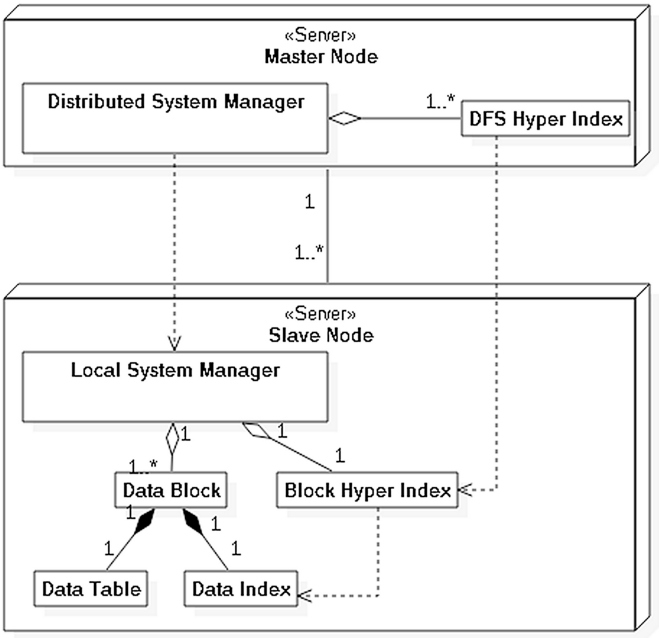


Fig. 1. Concept view of hierarchical indexes in HIM

Index is a high level index which keeps the hot data index collected from Block Hyper Index. Similar like Hadoop, real data are still stored in slave nodes organized in data blocks. Data blocks in HIM organize original data in table structure and generate index for each data table additionally. As a result, data blocks in HIM are dynamically changed based on user requirement. Each slave node also maintains a block hyper index which contains all hot data index collected from data index inner data blocks.

Integrating the hierarchical index based data storing mechanism with Hadoop's MapReduce data processing framework, our proposed HIM is able to do data analysis works more efficiently than the original Hadoop.

2.2 HIM-PRS: HIM Based Patent Recommendation System

On top of HIM, we built a Patent recommendation system which is designed for high speed patent searching and semantic patent recommendation. Instead of using 2 type of servers, HIM-PRS requires an additional control server for handling the entire patent analysis and service processing flow.

Our proposed system is designed to interact with a global scale knowledge base provided by linked data project. Therefore, SPARQL is the default knowledge query language used by HIM-PRS for communicating with remote knowledge node.

As Fig. 2 shows the HIM-PRS architecture. HIM-PRS control node maintains all modules related with our patent services. To provide our service, firstly, Patent Feature

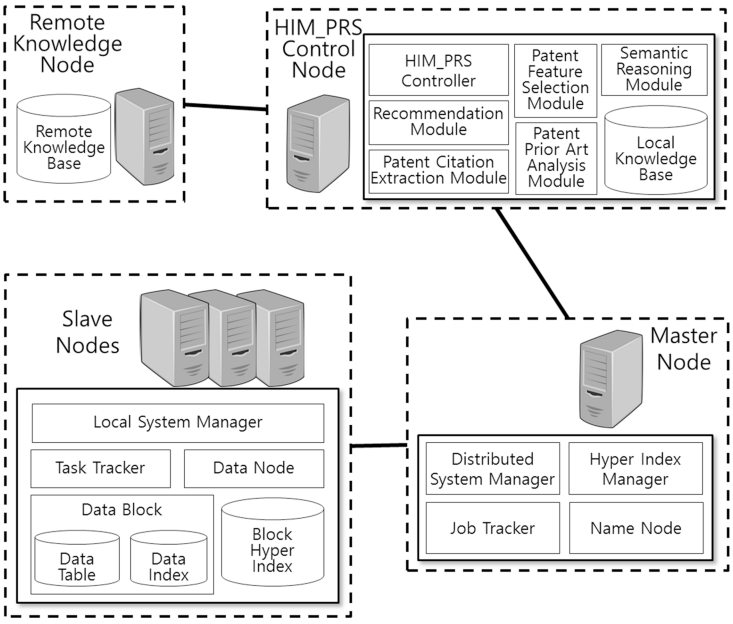


Fig. 2. HIM-PRS architecture

Select Module and Patent Citation Extraction Module analyze patents and extract necessary information such as patent feature words and citation data. Then, Patent Prior Analysis Module will generate patent prior art graph based on patent citation information. These 3 modules are pre-processing modules. Secondly, Recommendation Module receives keywords from users and transfers those keywords to Semantic Reasoning Module. The Semantic Reasoning Module first queries local knowledge base via SPARQL, and if there is no result found, it directly requests to remote knowledge base via the same SPARQL query.

3 Experimental Results and Analysis

In this section we will show a performance test of HIM and a patent recommendation example. We utilize total 14 intel dual core machines for running our experiment. Machines equipped with 2 GB memory work as HIM-PRS Control Node and HIM Master Node respectively. And the rest machines work as slave nodes. All machines are connected through a 100Mbit switch. Approximate 106 GB patent data which includes 1328892 patent documents collected from Google bulk data have been uploaded to our HIM-PRS before our experiments.

Figure 3 shows the query time of different block size. The blue line represents query time of using non-indexed data block while the red line represents using indexed date block. Obviously, as the data size growth, the query time for data in non-indexed blocks increases linearly. In the other side, using indexed blocks the query time

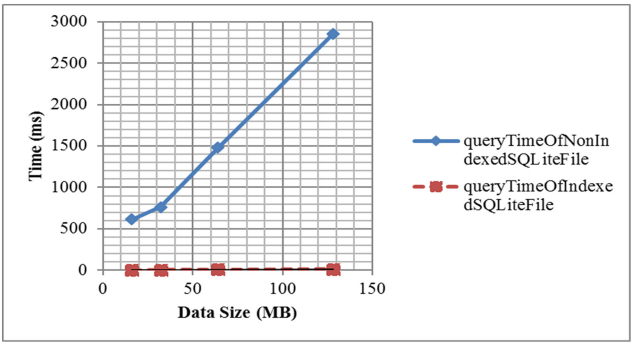


Fig. 3. Data query time for different block size

increases slightly. The non-indexed approach represents the original Hadoop solution. Our proposed HIM approach performs a better data query performance than the original Hadoop approach. Considering time for job preparation, jobs launching on top of HIM perform at least 2 times faster than original Hadoop. On top of HIM testbed, we built a HIM-PRS prototype. Currently we utilize WordNet [19] as our local knowledge base and DBpedia [20] as our remote knowledge base. In our example, we try "cloud computing" as our input key word.

Our proposed system utilizes a SPARQL query to query word "cloud" from local knowledge base WordNet. Similarly, we query the word "computing" also. Combined results of previous queries, we get a group of external keywords. Then, we send a query about "cloud computer" to remote knowledge base which is provided by DBpedia project through SPARQL. We get another group of semantic related words. By combining both groups of words, we have our final keywords. Finally, we searching patents which contain those keywords and ranking them based on our prior art analysis result. Based on system patents 7225249 and patent 8341462. The first patent claims a web based application which works under a PaaS pattern and the second patent claims a cloud provisioning technology. Both them are related with cloud computing topic and none of them shown in google patent searching page.

4 Conclusion

In this paper, we present a patent recommendation service system which is built on top of a performance enhanced Hadoop platform: HIM. Our proposed system performs a better patent data query performance. Meanwhile, our proposed system considers the semantic meaning of input keywords for patent recommendation.

We compare our platform with original Hadoop platform and google patent searching service. Our system shows a better data query performance than original Hadoop. However, if the query word is not indexed, our query will still as same as original Hadoop. In fact, if we build index for all data, the index size will be incredible. Thus to balance the size of index and performance of querying data is a new challenge for index based MapReduce platform and which is our future work.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the University Information Technology Research Center support program (IITP-2016-R2720-16-0004) supervised by the IITP (Institute for Information & communications Technology Promotion).

References

1. Patent Dispute Report (2015). http://www.unifiedpatents.com/news/2016/5/30/2015-patent-dispute-report
2. Apple Inc. v. Samsung Electronics Co. https://en.wikipedia.org/wiki/Apple_Inc._v._Samsung_Electronics_Co
3. USPTO. U.S. Patent Statistics Chart: Calendar Years 1963–2015. http://www.uspto.gov/web/offices/ac/ido/oeip/taf/us_stat.htm
4. USPTO. USPTO Patent Search, http://www.uspto.gov/patents/process/search/
5. Google. USPTO Bulk Downloads: Patents. https://www.google.com/googlebooks/uspto-patents.html
6. Google. Google Patent Search. https://www.google.co.kr/#tbm=pts&gws_rd=cr
7. Mancusi, M.L.: Technological specialization in industrial countries: patterns and dynamics. Weltwirtschaftliches Archiv. **137**(4), 593–621 (2001)
8. Narin, F., Carpenter, M.P., Woolf, P.: Technological performance assessments based on patents and patent citations. IEEE Trans. Eng. Manage. **1984**(4), 172–183 (1984)
9. Aurek. http://starwars.wikia.com/wiki/Aurek
10. Taduri, S., et al.: A patent system ontology for facilitating retrieval of patent related information. In: Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance. ACM (2012)
11. Dean, J., Ghemawat, S.: MapReduce: a flexible data processing tool. Commun. ACM **53**(1), 72–77 (2010)
12. White, T.: Hadoop: The Definitive Guide. O'Reilly Media Inc., Sebastopol (2012)
13. Abouzeid, A., et al.: HadoopDB: an architectural hybrid of MapReduce and DBMS technologies for analytical workloads. Proc. VLDB Endow. **2**(1), 922–933 (2009)
14. George, L.: HBase: The Definitive Guide. O'Reilly Media Inc., Sebastopol (2011)
15. Thusoo, A., et al.: Hive: a warehousing solution over a map-reduce framework. Proc. VLDB Endow. **2**(2), 1626–1629 (2009)
16. Dittrich, J., et al.: Hadoop ++: making a yellow elephant run like a cheetah (without it even noticing). Proc. VLDB Endow. **3**(1–2), 515–529 (2010)
17. Chang, F., et al.: Bigtable: a distributed storage system for structured data. ACM Trans. Comput. Syst. (TOCS) **26**(2), 4 (2008)
18. Rui, X., Kim, B., Min, D.: An efficient patent storing mechanism based on SQLite on Hadoop platform. In: Chen, Y., Balke, W.-T., Xu, J., Xu, W., Jin, P., Lin, X., Tang, T., Hwang, E. (eds.) WAIM 2014. LNCS, vol. 8597, pp. 382–392. Springer, Heidelberg (2014). doi:10.1007/978-3-319-11538-2_35
19. Miller, G.A.: WordNet: a lexical database for English. Commun. ACM **38**(11), 39–41 (1995)
20. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC/ISWC -2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). doi:10.1007/978-3-540-76298-0_52

Finding Meaningful Chronological Pattern of Key Words in Computer Science Bibliography

Joo-Seong Heo, Hyun-Kyo Lim, Kyong-Han Kim, and Youn-Hee Han(✉)

Laboratory of Intelligent Networks, Advanced Technology Research Center, Korea University of Technology and Education, Cheonan, South Korea
{chill207, goslim56, glenn89, yhhan}@koreatech.ac.kr

Abstract. In order to find meaningful patterns of research trends in the computer science and engineering field, we crawl a significant amount of bibliographic information, 3 million or more scholarly papers published from 1956 to 2015. We also make a list of target key words and analyze their frequency rate over the past 60 years based on the terms extracted from the titles and abstracts of the papers. We apply *k*-means clustering analysis for the target key words, and present a meaningful chronological pattern of target key words in the computer science engineering field over the past 60 years.

Keywords: Text mining · k-means clustering · DBLP

1 Introduction

Identification of meaningful patterns of research trends and extraction of potential knowledge from large volumes of text data are important tasks in various fields [1, 2]. In particular, the advent of high-speed internet generates large amounts of textual data in a variety of forms. However, it is difficult to analyze text data since there are irregularity and complexity in terms of textural data structure. Text mining is the process of deriving useful and meaningful information from text. Unlike conventional data mining tasks that extract the pattern from structured data sets, text mining is intended to explore relationship among the objects stored in unstructured data sets. In general, text mining involves several steps: (1) construction of the structured database from unstructured text inputs, (2) extraction of patterns and trends from the structured data, and (3) evaluation and interpretation of the patterns and trends.

A number of studies have been conducted to extract implicit information from large volumes of text data. Recently, some researchers have attempted to identify emerging technology trend from a large volumes of patent documents [3, 4]. Identifying research trend from scholarly papers has been attempted [5–8]. In particular, the authors in [5, 6]

Following are results of a study on the "Leades INdustry-university Cooperation" Project, supported by the Ministry of Education, Science & Technology (MEST).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_131

applied clustering analysis as well as a low-dimensional embedding method, association rule, and social network analysis to find meaningful associative patterns of the key words frequently appeared in the papers. However, they crawling and analysis were performed on just thousands of papers published in one journal.

In this paper, we attempt to identify research trends in the computer science engineering area using *k*-means clustering analysis. In order to that, we collect 3 million or more paper's metadata from the DBLP database [9]. DBLP is a representative bibliography database in the field of computer science and engineering. The publication years of the 3 million or more papers are distributed across the past 60 years. We also make a list of target key words and analyze their frequency rate over the past 60 years based on the terms extracted from the titles and abstracts of the papers. Based on our clustering analysis, we identify meaningful chronological pattern of key words and provide an understanding of research trends in the computer science engineering area.

The rest of the paper is organized as follows. Section 2 presents the data collection process and a list of keywords selected. Section 3 presents the clustering methods used in the study. Section 4 presents our analysis results. Finally, Sect. 5 gives some concluding remarks.

2 Data Crawling and Collection

DBLP is one of the representative database in the field of computer science engineering. We extracted the titles and abstracts from each of 3 million papers published from 1956 to 2015. For our analysis, we also made a list of 2896 IT terms based on Gartner IT Term [10]. Among such IT terms, we selected 1397 important terms frequently appeared in the titles and abstracts of the crawled papers, and we call them target key words. Table 1 shows that the number of all IT terms and target key words used for our analysis, and Fig. 1 presents a list of selected target key words.

Table 1. Number of terms used for our analysis

	All IT terms	Target key words
Numbers	2896	1397

network, internet protocol, web, openstack, nfv, deep learning, protocol, framework, tcp, hypertext, programming language, synchronization, terminal, data center, mobile ip, database design, social network analysis, ftp, bandwidth, buffer, streaming, operating system, clustering, scalability, html5, web crawler, microprocessor, artificial intelligence, ipv6, c++

Fig. 1. A list of selected target key words

In the DBLP database, we can find 3 million or more papers published from 1935 to 2015. We divide them into 8 groups, each of which contains the papers published for eight decades: 1935 to 1945, 1946 to 1955, 1956 to 1965, …, 2006 to 2015. However, there are 113 and 398 papers in the first decade '1935 to 1945' and the second decade '1946 to 1955', respectively, which are much fewer than the number of papers published for other decades. Therefore, we use only paper information generated over the past six decades (i.e., 60 years).

Based on the target key words and the list of paper group per decade, Table 2 represents the frequency of the target key words over the six decades. In the table, for example, we can find the term 'network' occurs 69 times in the title or abstract of the papers published during 1956 to 1965 (D1960) and 'openstack' 160 times during 2006 to 2015 (D2010). We use the Table 2 dataset to identify research trends over the past 60 years.

Table 2. The dataset containing the frequency of the keywords in period

Six decades	D1960	D1970	D1980	D1990	D2000	D2010
	1956–1965	1966–1975	1976–1985	1986–1995	1996–2005	2006–2015
Number of papers	5096	22827	66566	248179	795146	2113389
Network	69	337	1763	9848	42465	151642
Openstack	0	0	0	0	0	160
…	…	…	…	…	…	…
Digital	199	481	1219	3243	15604	35176
Data center	0	2	16	26	132	4043

3 Clustering Analysis

In our analysis, *k*-Means clustering partitions n target key words into k clusters in which each key word belong to cluster with the nearest mean. The brief summary of our k-means clustering method is as follows:

(1) Given k initial points that can arbitrarily determine, each key word is assigned to one of the k initial points close to the key word, creating k clusters.
(2) These initial points are then replaced with the mean of key words currently assigned in the clusters.
(3) This procedure is repeated with updated points until assignments do not change.

The outcome of the *k*-means clustering depends on the number of clusters (k) and the distance (or similarity) metrics. In our analysis, the distance d between two target key words, w_1 and w_2, is defined as follows:

$$d = \sum_{i=1}^{6} \left| f_i^1 - f_i^2 \right| \tag{1}$$

where f_i^1 and f_i^2 are the frequency of the target key word w_1 and w_2 in the ith decade, respectively.

4 Analysis Results

As our analysis software tools, we used MySQL 5.5.49 on the Linux-ubuntu.14.04.1, pyspark equipped with spark-1.6.0, and mysql-connector-java-5.1.38. In our analysis, *k*-means clustering was performed on the dataset containing the frequency of the target key words over six decades shown in Table 2. We set $k = 8$ and used Eq. (1) for the distance metric between two target key words. Figure 2 shows the frequency rate profiles of eight clusters from *k*-means clustering analysis. We classified the eight clusters into 4 groups, 'Steadily Increase & Until Recently', 'Steadily Increase but Recently Decrease', 'Once Popular', and 'Mostly Decrease', according to the chronological pattern of their frequency change. For examples, the frequency rate of target key words in 'Cluster 1' steadily decreases across six decades. On the other hand, the frequency of those in 'Cluster 2 and 8' tends to increase until D1980 but decrease rapidly over the past 30 years. Also, the frequency of those in 'Cluster 5 and 6' tends to increase steadily over five decades but decrease recently over the past 10 years, and we can finally find the frequency of those in 'Cluster 3, 4 and 7' increases steadily and also even recently over the past 60 years.

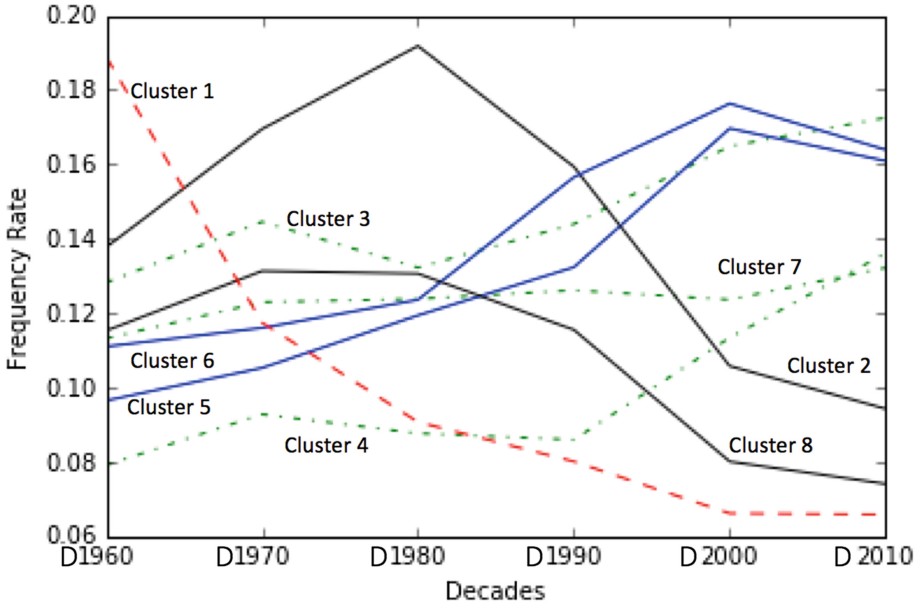


Fig. 2. The frequency rate profiles of nine clusters of the target key words observed over the past 60 year (1956–2015)

A list of the selected target key words that belongs to each group is shown in Table 3. This result can provide an understanding of research trends in the computer science engineering field over the past 60 years.

Table 3. A list of selected key words in each of three group based on their patterns.

Group	k-means clusters	Selected target key words	Description
Group 1	Cluster 3 Cluster 4 Cluster 7	Access point, active directory, adaptive learning, affective computing, analytic applications, analytics, application architecture, authentication technologies, autonomous system, autonomous vehicles, backbone, backbone network, bandwidth, base station, batch processing, benchmarking, bioinformatics, business analytics, content management, concurrent engineering, cognitive radio, clustering, cloud computing, circuit switching, checksum, code division multiple access, design thinking, deep learning, data scientist, data center, distributed data management, dynamic routing, eye tracking, femtocells, fraud detection, grid computing, html5, internet protocol, ipv6, job scheduling, media access control, mobile ip, mobile network, mobile social networks, multithreading, network intelligence, network management, network virtualization, openstack, p2p, protocol stack, proxy servers, saas, sdh/sonnet, session initiation protocol, social analytics, social network analysis, software defined networking, switched network, tcp/ip, text analytics, virtual reality, virtualization, vlan, vpn, web crawler, wireless broadband	Steadily increase up to recently
Group 2	Cluster 5 Cluster 6	Architecture, availability, broadcast, buffer, cache, clock, community, data mining, encryption, framework, fuzzy logic, information technology, java, load balancing, machine learning, mobile, protocol, qos, scalability, signature, streaming, tcp, throughput, visualization, web, web services, xml	Steadily increase but decrease recently
Group 3	Cluster 2 Cluster 8	Artificial intelligence, attenuation, broadband, c++, channel capacity, circuit board, composition, cycle time, database design, database management system, distributed computing, distributed database, expert system, hypertext, information management, knowledge representation, microprocessor, operation system, packet switching, parallel processing, response time, software	Once popular

(*continued*)

Table 3. (*continued*)

Group	k-means clusters	Selected target key words	Description
		development, speech recognition, synchronization, terminal, transducer, workstations	
Group 4	Cluster 1	Binary code, modulation, redundancy, switch, control, application, information, processing, simulation	Steadily decrease

5 Conclusions

In this paper, we revealed research trends of the computer science engineering field by providing the chronological pattern of key word frequency change. We employed *k*-means clustering analysis to draw out the meaningful patterns and research trends over the past 60 years. In the future, we will also study another analysis method (e.g. social network analysis, locally linear embedding, and association rule) based on key words frequently appeared in the papers.

References

1. Ronen, F., Sanger, J.: The Text Mining Handbook - Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, Cambridge (2007). DAGLIB
2. Kao, A., Poteet, S.R.: Grid Natural Language Processing and Text Mining. Springer, London (2007)
3. Tsenga, Y.-H., Linb, C.-J., Linc, Y.-I.: Text mining techniques for patent analysis. Inf. Process. Manage. **43**(5), 1216–1247 (2007)
4. Lee, S., Lee, S., Seol, H.: Using patent information for designing new product and technology: keyword based technology roadmapping. R&D Manage. **38**(2), 169–188 (2008)
5. Cho, S.-G., Kim, S.-B.: Identification of research patterns and trends through text mining. Int. J. Inf. Educ. Technol. **2**(3), 233–235 (2012)
6. Cho, S.-G., Kim, S.-B.: Finding meaningful pattern of key words in IIE transactions using text mining. J. Korean Inst. Ind. Eng. **38**(1), 67–73 (2012)
7. Poelmans, J., Ignatov, D.I., Viaene, S., Dedene, G., Kuznetsov, S.O.: Text mining scientific papers: a survey on FCA-Based information retrieval research. In: Perner, P. (ed.) ICDM 2012. LNCS (LNAI), vol. 7377, pp. 273–287. Springer, Heidelberg (2012). doi:10.1007/978-3-642-31488-9_22
8. Jomsri, P., Prangchumpol, D.: A hybrid model ranking search result for research paper searching on social bookmarking. In: Proceedings of INISCom 2015, pp. 38–43 (2015)
9. DBLP. http://dblp.uni-trier.de
10. Gartner IT term. http://blogs.gartner.com/it-glossary

The Design of Intelligent Video Analytics System Performing Automatic Noise Rejection by Comparing Distribution of Metadata of Moving Object

Taewoo Kim(✉), Hyungheon Kim, and Pyeongkang Kim

Technology Laboratory, Innodep Inc., Digital, 31, Guro-gu, Seoul 08375, Korea
{davidkim, josephkim, david}@innodep.com

Abstract. For growing interest in public safety in Korea, the number of CCTV to be monitored is rapidly increasing. It's impossible to monitor all CCTV cameras with limited monitoring agents. So, computer video analysis solution is essential. A lot of research concerning this demand is conducted but it's not enough for them to analyze lots of videos in integrated control centers because most of them use heavy algorithm for analysis. For this reason, we implemented an intelligent video analytics system which employs engine with light algorithm and data analysis of metadata from that engine to analyze thousands of CCTV videos for detecting object of interest. Also, we introduce technique for automatically rejecting noise object from that engine.

Keywords: CCTV · Moving object · Metadata filtering · Noise rejection

1 Introduction

It is expected that the number of CCTV in Korea will increase in the ratio of 20 % annually from 560,000 in late 2013 to 700,000 in 2015. The number of CCTV is much greater if you include private CCTV also, which is 5,000,000, and one is being captured by those CCTVs every 14.2 m and 5.5 s [1, 2]. However, the number of monitoring agents hasn't grown in proportion to the rate. As a result, one should monitor lots of cameras, which leads to low concentration and poor effectiveness of agents.

One study [3] conducts experiments regarding effects of multiple camera monitoring upon agents. It showed that one with 9 cameras lost about 60 % of events and another with 4 cameras lost about 20 % of events. This outcome indicates that introductions of computer-aided video analysis are essential. To meet that necessity, lots of research has been carried out. Generally, video analytics department consists of followings, which are optimum camera placement, data acquisition and storage, feature extraction, object detection and tracking, interpretation of video data, data visualization, and analytics algorithms [4]. Each research department pursues indigenous goals.

Those video analysis techniques were studied to be applied to commercials. Among them are research on recognition algorithms for human behavior in transit scenes [5] and methodology for application of intelligent video systems to integrated control centers [6]. Specifically, for use in integrated control centers, because of their focus on

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_132

moving objects such as human, cars and their weird and trouble making action, studies for recognition of those high semantic actions has been conducted, such as fighting, crime, car accidents and etc. [7].

Aforementioned studies are for detecting and recognizing objects very specifically and precisely and they use very heavy algorithms. With that engines with heavy algorithms, it's not possible to analyze thousands of videos in integrated control centers. So, in this study, we decided to employ light engine and algorithm for analysis. Instead we did in-depth analysis of metadata of moving objects, which are from the video analytics engine we made. And we made a system which filters out environment noise utilizing that analysis result and recommends objects of interest.

In this paper, we describe the components of our system and techniques used in chapter 2. In chapter 3, we implement each components of the system and analyze the results. In chapter 4, conclusions and directions for further studies are presented.

2 Design of Intelligent System for Target of Interest

The total system consists of two main marts. First part performs video analysis and saves metadata of moving objects. Video analysis separates foreground for detecting any moving objects and does blob analysis upon them. Second part does data analytics on metadata from the former part. With the analysis result, it filters out objects considered to be noise and shows other objects to monitoring agents. Meanwhile, noise filtering is done by plotting 2D scatter plot of dwell time, total distance of each object and exempt same pattern among several time periods. The whole system configuration is as follows (Fig. 1).

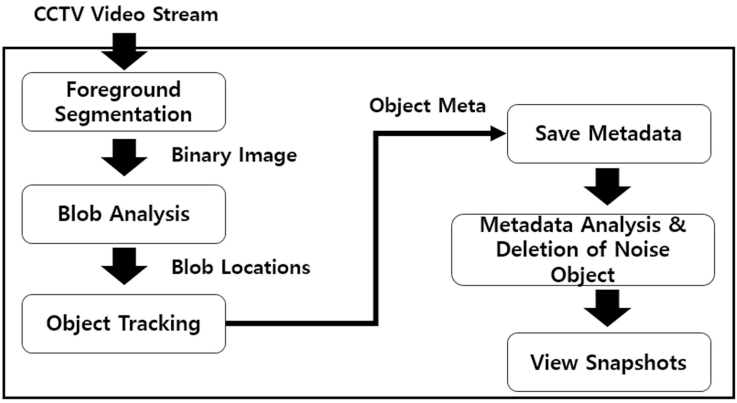


Fig. 1. The configuration of intelligent video analytics system.

2.1 Foreground Segmentation

In this system, background is modeled using Gaussian mixture model and each pixels are divided into foreground or background. Foreground segmentation comprises two

processes. First, background image pixels are modeled with k individual Gaussian variables and changed pixels are recognized per each frame. Second, depending on computing resources, model parameters are updated constantly. By doing this, the background models adapts to changing light conditions over time.

2.2 Blob Analysis

This part does blob analysis with the foreground pixels data from the former part to recognize pixel groups as objects. And it only clusters pixel groups if the size of each group exceeds some level to suppress noise. The level is determined considering the size of frames.

2.3 Object Tracking

Using object locations from the blob analysis, tracking is performed. Depending on the location estimation of Kalman filter, each object in each frame is tested if it belongs to any object track existent. If the location of recognized object is in existent tracks, the new position of it is recorded in track information metadata. If not, it is considered as a newly detected object and given a new track ID.

2.4 Save Metadata of Objects

The information about object, which is from the former parts, is all saved. Those are called metadata of objects. We describe each elements in the metadata in Table 1.

Table 1. Metadata components list and their type

Component	Description
Object ID	Unique number assigned to each newly detected objects.
Thumbnail image	Representative image for an object.
Moving track	M by 5 matrix. Column from 1 to 5 indicates local time, object's x coordinate, object's y coordinate, width and height respectively. M increases as the time passes.
Appearance location	The location coordinate of object when it first appeared.
Disappearance location	The location coordinate of object when it disappeared.
Maximum blob size	The maximum size of object.
Dwell time	The existence duration of object.

2.5 Metadata Analysis and Deletion of Noise Objects

Generally, it's likely that the locations of environment noise doesn't change much and their durations are short. Useful properties for deleting environment noise by making use of those characteristics are dwell time and moving track in metadata in Table 1. By ruling out objects under dwell time threshold and under total distance threshold, noise objects are deleted and objects of interest are extracted. By the way, the total distance is acquired from moving track information. To visualize distribution of metadata of objects, we draw 2D scatter plot per each object, whose X axis is dwell time of each object and whose Y axis is total distance. The metadata collection period is 5 min. This is shown in Fig. 2.

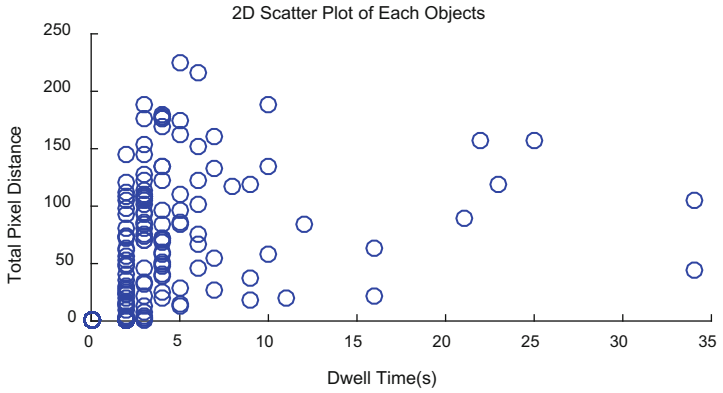


Fig. 2. 2D scatter plot of objects, whose x axis is dwell time and y axis is total distance.

In this figure, the objects of interest are in the upper right half plane. They randomly occurs and their distribution is dispersive. Meanwhile, noise objects are in lower left half plane. Their population is extremely large and repetitive. So, using this fact, we discriminate noise elements by finding similar distribution over time. For this purpose, we made m by n grid upon 2D scatter plot and counted objects in each grid bin for 5 min. We made several 5 min sets and correlated with each other. The bins with high correlation value are treated as environment noise because the large correlation value means the same pattern over time and it's likely for the bins to contain environment noise objects.

3 Analysis of Implemented System

We implemented each component in the designed system in chapter 2 using matlab for video analysis. Figure 3 shows the result of foreground segmentation on the left and blob analysis result on the right. For the system uses light algorithm for simple object detection, there were lots of noise objects detected.

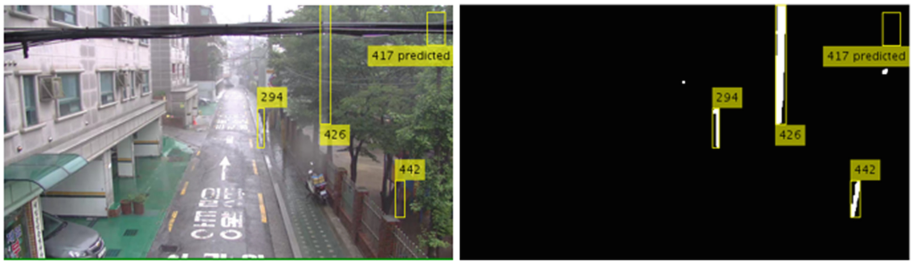


Fig. 3. Foreground segmentation & blob analysis result.

Among whole metadata, we made five sets of metadata. Each set is metadata for 5 min, so five sets account for 25 min. Using each set, 2D scatter plot could be acquired like in Fig. 2. Among the sets, we correlated two sets, which is done by multiplying each component in the same bin. The Fig. 4 shows the result of correlation.

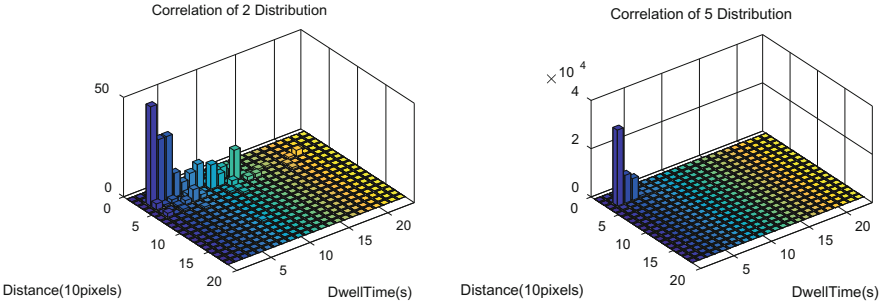


Fig. 4. Correlation result of 2(left) and 5(right).

As the number of sets increases, the value of bins in possible noise location also increases. This is because the noise occurs repetitively in a similar manner over time. On the other hand, the correlation value of object of interest diminished. This is because of their randomly occurring properties. The snapshots of objects after deletion of environment noise are presented in Fig. 5.

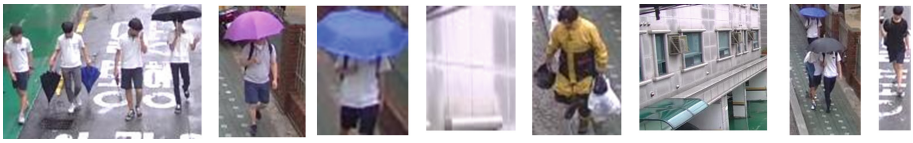


Fig. 5. Snapshots of objects of interest.

As you see in those pictures, it showed reliable results for detecting objects of interest, though there were some noise objects in the list.

4 Conclusion

Currently, more and more safety threats are emerging in Korea. So, to suppress them, the number of CCTV is also increasing day and day. Therefore, to keep an eye on those lots of cameras, computer analysis technology is a compulsive need. In this paper, we implemented an intelligent system for automatically detecting target of interest and rejecting noise objects using a light engine and heavy data analysis of object metadata. It is expected for this system to improve efficiency of monitoring agents by analyzing lots of CCTV simultaneously. Whereas, we expect further research on techniques and algorithms for detecting more specific events which are useful to monitoring agents, such as fighting, crime and car accidents.

Acknowledgments. This work was supported by the ICT R&D program of MSIP/IITP. [B0101-15-0266], Development of High Performance Visual BigData Discovery Platform for Large-Scale Realtime Data Analysis.

References

1. Statistics Korea. http://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=2855
2. Hello.net News. http://www.hellot.net/new_hellot/magazine/magazine_read.html?code=202&sub=003&idx=24743
3. Noah, S., Thomas, S., Dmitry, G., Rangachar, K.: How effective is human video surveillance performance. In: 19th ICPR International Conference on Pattern Recognition, pp. 1–3. ICPR, Tampa (2008)
4. Ayesha, C., Santanu, C.: Video analytics revisited. IET J. Comput. Vis. **10**, 237–247 (2016)
5. Joshua, C., Matthew, S., Dmitry, B.G., Deborah, B.S., Rangachar, K.: Understanding transit scenes: a survey on human behavior. IEEE Trans. Intell. Transp. Syst. **11**, 206–224 (2010)
6. Honghai, L., Shengyong, C.: Intelligent video systems and analytics: a survey. IEEE Trans. Ind. Inf. **9**, 1222–1233 (2013)
7. Subetha, T., Chitrakala, S.: A Survey on human activity recognition from videos. In: 2016 International Conference on Information Communication and Embedded Systems, pp. 1–7. ICICES, Chennai (2016)

Dependability Analysis of Digital Library Cloud Services with Load Sharing

Dongseok Lee, Sungsoo Kim(✉), and Tae-Sun Chung

Computer Engineering, Ajou University,
Suwon, Republic of Korea
{ehd0019, sskim, tschung}@ajou.ac.kr

Abstract. Cloud service utilizes computing resources to store all of information on the Internet. It means that we can use the service anytime, in anywhere through a variety of IT equipment. For reliable operations of digital library cloud services, it is necessary to ensure dependability. In this paper, we apply the cold standby, hot standby, and load sharing in the system and compare our system with existing systems to measure the availability and reliability of dependability characteristic.

Keywords: Dependability · Availability · Reliability · Cloud computing · Redundancy strategy · Load sharing

1 Introduction

Cloud computing is an intangible form like clouds and it stores all of information from the user to the server on the Internet. And it includes an environment which ensures high availability in any time and in anywhere through various IT equipment. In a cloud computing environment, we can store data safely because data is stored on external servers. For example, Digital Library Service has an advantage due to its high connectivity at anywhere.

Digital library is a collection of data stored on various media such as tapes, books, publications, PC, videos, and disks. Then, organizations/universities come up with Digital Library Service with their own materials for people to gain knowledge from it.

Due to the expanding growth of information Digital Library has, dependability becomes more important. If the service is interrupted by an unexpected error, it can cause serious damages causing data loss/network failure. Furthermore, it is necessary to spend additional costs for data recovery, in some cases, may not be able to fully restore the data. Thus, in order to provide a consistent and reliable digital library service, it is necessary to study scheme that can improve the dependability characteristics like availability or reliability.

Many researchers have measured the availability applying hot standby/cold standby in a cloud service environment [1]. However, considering that the application fields of cloud computing have been getting wider with increased users' needs, the existing researches for the dependability are remarkably insufficient to fulfill user's desired conditions. In this paper, we propose a load sharing strategy to enhance the dependability

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_133

and measure the availability and reliability by comparing it with redundancy of conventional systems.

2 Models

A. Dependability

In recent years, dependability has emerged as major interest in various fields which requires construction of computing systems. Especially in particular areas where failures in the system can cause fatal problems, the system should possess the dependability. Dependability comprises availability, reliability, safety, maintainability, confidentiality and integrity. In this paper, we analyze availability and reliability.

Availability is the degree to which a system, subsystem is in a specified operable and committable state. Availability of a system is typically measured as a factor of its reliability – as reliability increases, so does availability. It is important to increase the operable time of systems to increase the availability.

$$\text{Availability} = \frac{MTTF}{(MTTF + MTTR)} \tag{1}$$

Generally, the availability is represented by the Eq. (1), comprised of MTTF (Mean Time To Failure) and MTTR (Mean Time To Repair). Reliability means the probability of system device that performs the operations correctly without failure for a period of time. In the mathematical aspect, it can be represented by MTTF or possibility that failure has not occurred during a mission period. As a purpose of these properties, there are dependability methods such as fault prevention, fault tolerant, fault removal, fault forecasting and so forth. Fault tolerant is a technique placing the backup unit and utilizing it to maintain the functionality of the system against failure occurs. The redundancy strategy is one of fault tolerance methods.

As shown in Table 1, the redundancy strategy is composed of hot standby, cold standby, warm standby and load sharing. This method constitutes an extra module in the interior of the system. Even if errors occur, it provides stable operation by maintaining the operating state of the system.

There is a big difference between the standby redundancy and the load sharing system. In the standby mode, the standby module remains ready until the main module stopped working due to failures. In the load sharing, it processes works simultaneously without a distinction between the main module and the standby module. For that reason, load sharing outperforms standby mode in terms of throughput and speed. In this paper, we measure the level of availability and reliability comparing the load sharing strategy with the standby systems. In load sharing, load pattern is a constant and constant pattern load can be applied to a wide range of systems. Also, it applies the equal load sharing rule to each module running a constant load to increase processing speed [2].

Table 1. Redundancy strategy

Strategy		Attribute
Standby redundancy	Hot standby	Maintaining the operation state as the same level of the main module and retaining system functions Error rate during wait: Same as the main module
	Warm standby	Maintaining the operation state in lower level than the main module and retaining system functions Error rate during wait: greater than 0 and less than the main module
	Cold standby	Idle state Error rate during wait: 0
Load sharing		Structure is designed with an extra structure to share the work without distinguishing the main module

B. Reliability Block Diagram

Reliability Block Diagram (RBD) is a model represented by a block diagram to evaluate the reliability of the entire system. Each block will be described by using a system component failure rate and it is possible to calculate a different dependability characteristic such as availability and maintainability [3].

In Fig. 1, a Frontend Model is represented by using RDB. Frontend model, a management server is composed of Hw (Hardware), OS (Operational System) and Mng (Management Server). Figure 2 shows Node Model and it is composed of 5 block (Hw, OS, Mng, VM (Virtual Machine) and DL (Digital Library). The Node Model applies the redundant strategy. In conventional systems which connect the models shown in Figs. 1 and 2, if an error occurs in either part of them, the whole system can be stopped.

Fig. 1. Frontend model

Fig. 2. Node model

Figure 3 is a state of connected Node Model by applying redundancy. Using this model, we can compare the availability and reliability by applying the cold standby, hot standby and load sharing. And MTTF, MTTR of each module is applied by referencing Table 2.

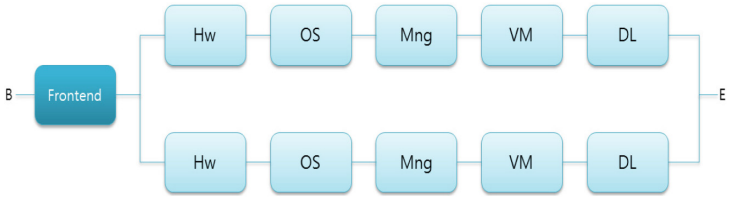


Fig. 3. Redundancy system model

Table 2. Input Parameters

Parameters	MTTF	MTTR
HW	8760 h	lOOmin
Management tool	788.4 h	lhr
OS	1440 h	lhr
VM	2880 h	lOmin
Digital Library	6865.3 h	lOmin

C. State transition diagram

Figure 4 shows the state transition diagram which applies the load sharing. Each node is a i.i.d (independent and identically distributed), followed by exponential life-time distribution. λ represents the failure rate and μ represents the repair rate. State transition diagram can be represented as states, working, failure and repair [4]. The Load Sharing system basically consists of a parallel system. First, we consider a parallel system which consists of two nodes. Assuming that it is possible to repair one node at a time, S_0 is a state in which two modules are operating properly, S_1 is a state where an error occurs in one of the modules, and S_2 is a state that both two modules are broken so it cannot perform any normal operation.

$$A_s = \prod_{j=1}^{y} \left(1 - \frac{\lambda_0..\lambda_{n_j-k_j}}{\mu_1..\mu_{n_j-k_j+1}} \left(1 + \sum_{i=1}^{n_j-k_j+1} \frac{\lambda_0..\lambda_{i-1}}{\mu_1..\mu_i}\right)^{-1}\right) \tag{2}$$

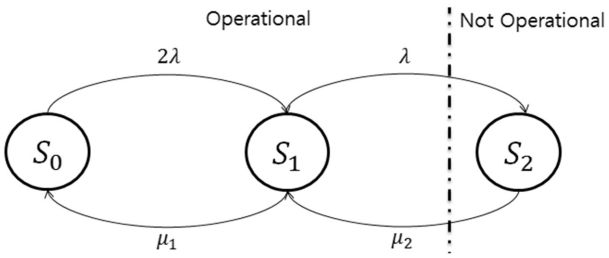


Fig. 4. Load sharing state transition diagram

$$R(t) = \exp(-2\lambda t)2[1 - \exp(-2\lambda t)]\exp(-\lambda t) \quad (3)$$

Equation (2) shows an availability formula of a system which applies the load sharing strategy [5]. A reliability formula of a system which applies the load sharing strategy is described in Eq. (3) [6].

3 Performance Analysis

Figure 5 shows the degree of availability after applying the redundancy to the Node model. Base is a basic series form and it has lower degree of availability than other results. Between hot standby and cold standby, the hot standby shows higher availability due to the faster operation time. Load sharing has higher availability than hot and cold standby because two nodes are in active state.

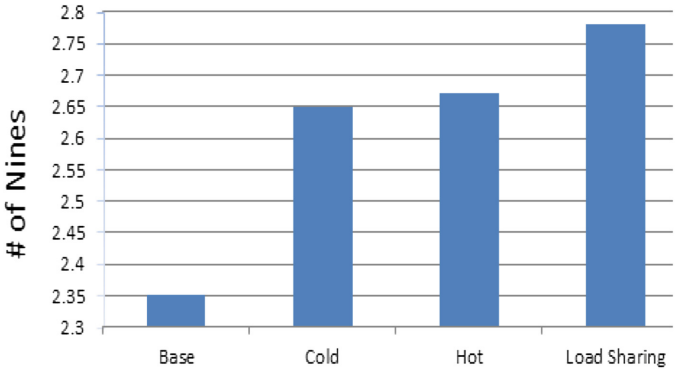


Fig. 5. Availability with comparison models

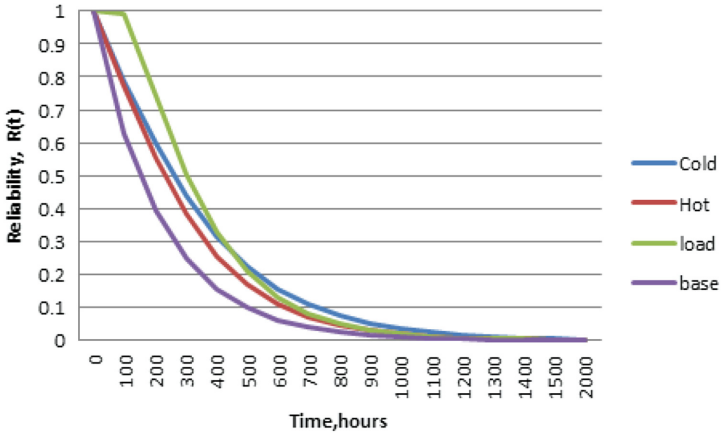


Fig. 6. Reliability result

Figure 6 shows the reliability according to the change of time. In case of basic series form, it shows that reliability decreases acutely than others. In case of hot standby, the reliability decreases more sharply than that of cold standby because it maintains the operating state at the same level as the main module. In the case of load sharing, initial reliability is measured higher than others, but as time passed reliability is rapidly lowered.

4 Conclusion

In this paper, we apply the Redundancy strategy to Digital Library in cloud computing environments to analyze the availability and reliability. Comparing with conventional systems, cold standby, hot standby and load sharing increase the availability for 12 %, 13 % and 18 % respectively. In terms of reliability, load sharing technique was measured that it outperforms other redundancy at only initial time. It shows that it would be solved by adding a component. Load sharing requires an additional cost, but is considered as the best solution for any system to improve dependability.

References

1. Araujo, J., Maciel, P.: Availability evaluation of digital library cloud services. In: 2014 44th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 666–671 (2014)
2. Amari, S.V., Misra, K.B., Pham, H.: Tampered failure rate load-sharing systems: status and perspectives. In: Handbook of Performability Engineering, pp. 291–308 (2008)
3. Dantas, J., Matos, R.: An availability model for eucalyptus platform: an analysis of warm-standby replication mechanism. In: 2012 IEEE International Conference on Systems, pp. 1664–1669, October 2012
4. Ng, S.L.: A fault tolerant load sharing replicated buffered banyan network. In: Algorithms and Architectures for Parallel Processing, ICAPP, pp. 726–735, April 1995
5. Arabi, A.A.Y., Jahromi, A.E.: Availability optimization of a series system with multiple repairable load sharing subsystems considering redundancy and repair facility allocation. Int. J. Syst. Assur. Eng. Manag. **4**(3), 262–274 (2013)
6. Xie, M., Dai, Y-S.: Computing System Reliability: Models and Analysis, pp. 58–66 (2004)

Document Classification Using Word2Vec and Chi-square on Apache Spark

Mijin Choi, Rize Jin, and Tae-Sun Chung(✉)

Computer Engineering, Ajou University, Suwon, Republic of Korea
{chyocolate, rizejin, tschung}@ajou.ac.kr

Abstract. Text mining is a mechanism to find information by extracting resources from natural language. Compared with structured data in databases, text is unstructured and difficult to be dealt with for analyzing. Additionally, it is tedious tasks for users to identify accurate data. Text mining algorithm is similar to data mining, except that it processes data in database and aims to determine whether any document belongs to a specific topic. There are some classification algorithms. To identify which classifier is efficient, we compare SVM (Support Vector Machine) and Naïve Bayes, and use Apache Spark which is distributed system environment, to classify a large number of documents efficiently.

Keywords: Text classification · Naïve Bayes · SVM · Word2Vec · Chi-square statistics · Apache Spark

1 Introduction

Text mining is a mechanism of discovering information by extracting resources from natural language text. Compared with structured data in database, text is unstructured and difficult to be dealt with for analyzing. Additionally, it is tedious tasks for users to identify accurate data from natural language. Text mining is a variation on a field called data mining, which tries to find patterns. It is similar to data mining, except that it processes data in databases [1].

As data mining finds patterns in data, text mining finds patterns in text. However, there are differences between two concepts [1]. Data mining tools [2] are designed to handle structured data from databases, but text mining work on unstructured data such as emails, full-text documents. Also, data mining may be described as the extraction of implicit, unknown, and useful information from data. Information obtained by data mining designs is hidden and difficult to be extracted. However, with text mining designs, the information is clearly and explicit [1, 2].

But it is not easy to acquire accurate information from users. Classification algorithm helps people to understand and recognize characteristics of data because people often make mistake during analysis when they establish relationships between various features. Document categorization is a technique to automatically assign the document to a category which is defined in advance. It aims to reduce the large amount of job, manage documents and retrieve efficiently.

Document categorization process is composed of feature extraction process and document text classification process. For feature extraction, chi-square statistics [3], one

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_134

of the feature selections, is used. The chi-square has a problem. Since the weight of word is based on the number of appearance of the word, it only considers the characteristic of the word. That's why we use Word2Vec and compare Word2Vec with chi-square [4, 5]. Word2Vec is considered the characteristic of word as well as meaning of word. Word2vec has improved accuracy in the field of natural language processing for efficiently estimating the meaning of the word on the vector space and it is a technique for learning a characterization vector of the word through a neural network.

Naive Bayes classifiers and Support Vector Machine (SVM) belong to the supervised classification, which is popular when data is categorized. Naïve Bayes is one of the most widely used algorithms in the method of classifying a document. This design calculates the probability of the class defined within the categories belonging to specific categories depending on the feature with the document. Support Vector Machine (SVM) is used to analyze data and divide data into binary. Given a data to be classified into two classes, a SVM model assigns the data into one of the categories [6]. The purpose of the constructs the *hyperplane* for classifying the two classes based on structural risk minimization techniques.

Apache Spark [7] is the big data processing platform based on open source. It solves the problem of the delay of disadvantages due to disk I/O of Hadoop by storing the result of Mapreduce [8] operation in the main memory. Also, Spark also offers a tolerance by proposing a model called RDD (Resilient Distributed Dataset). Since RDD offers the function of recovery for data stored in-memory, Apache Spark can deal with data without hard disk.

As mentioned, Chi-square is less accurate than Word2Vec because weighting the words statistically. Word2Vec is considered the characteristic of word as well as meaning of word by vectorization. We compare the accuracy of Chi-square with Word2Vec. Also, by using Apache Spark, when handling a mass of the document, we try to find out Apache Spark's performance.

Our rest sections are organized as follows: describe the overview of relative algorism in Sect. 2, in Sect. 3, describe related word, in Sect. 4, and describe performance evaluation and in Sect. 5, conclusion.

2 Overview of Classification Algorithm

Classification is one of the data mining methods, which lets unstructured data make structured data. It helps people to discover knowledge and make decision. There are two steps in classification. First, a complex and huge training datasets are collected. At that time, "brute-force" is the simplest available measurement and analysis is done, and then patterns are created. Secondly, as shown in Fig. 1, to identify the accuracy of classification pattern, evaluation of datasets takes place [6].

2.1 Support Vector Machine (SVM)

SVM is a learning method introduced in order to address a pattern recognition problem of the two categories. Machine learning is the method that enables computers to learn

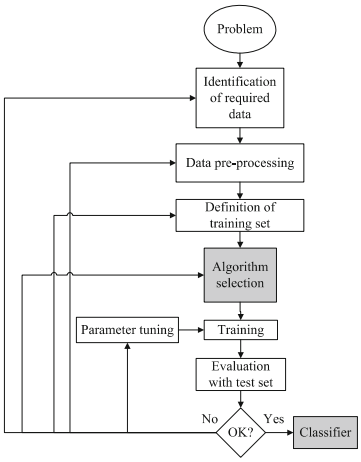


Fig. 1. Process of classification [6]

using algorithms and it is expanded into the prediction of behavior. As shown in Fig. 2 basic principle of SVM is that N points consisting of two categories are divided into one category. Although there are many boundaries that separates the two categories, SVM divides two groups through a *Hyperplane* which maximizes a *margin* that is determined by the certain points called *support vector*. The data which exists the nearest from *Hyperplane* is *support vector*. Support vector is important information that decides the *Hyperplane*.

Hyperplane can be described by equations.

$$y = wx + b \tag{1}$$

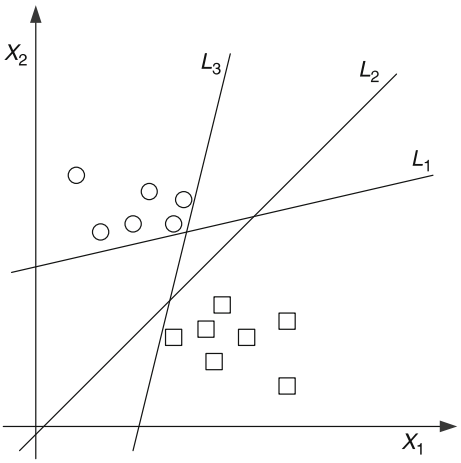


Fig. 2. Graph of SVM

The value w and b are acquired by learning but, x is the vector that separates attributes in two categories. Given attributes are divided 'A' or 'B', after learning the SVM. If $y = wx + b > 0$, the new object belongs to 'A', if $y = wx < 0$, it into 'B'. There are some advantages in SVM. First, SVM can deal with large volume of data because over-fitting protection is applied to SVM. Second, SVM is well useful for problems with dense and sparse attributes for each document. Finally, the idea of SVM is to find linear separators. SVM shows high performance. However, it has problems related to scalability due to time-consuming for scanning of big-volume datasets.

2.2 Naïve Bayes Classification Algorithm

Classifying documents means that it classifies the documents one category or multiple categories. In learning machine, the document classification algorithm learns features of the training document dataset while analyzing. It predicts and considers which category the new document belongs to by using the learned information.

Naïve Bayes classification algorithm is one of the most useful algorithms. It is a model based on Bayes theorem assuming that features are independent each other. Naïve Bayes Classifier is very simple but, popular as a traditional classification method. This methodology defines the class in category and calculates the probability belonging to a particular category depending on the feature of the document. Then, it classifies the document into one category representing the highest probability. The basic formula is as following.

$$P(C|F) = \frac{P(F) \times P(F|C)}{P(C)} \tag{2}$$

C refers to the class belonging to the category and F represents the feature that document has respectively. $P(C|F)$ is the probability that the document to be included in the class C when the document has the feature of F. Since $P(C|F)$ is objective output but, cannot be calculated directly, $\frac{P(F)\times P(F|C)}{P(C)}$ is transformed. $P(F)$ has the same value for all of classes as the probability of the feature on the all doc. $P(C)$ is the ratio belong to the class C of the all of documents so, $P(F|C)$ is the probability appearing in the C class F features.

3 Related Work

3.1 χ^2 Statistics

Feature selection is the way to reduce the amount of calculations that occur in the classification process by reducing the number of features. The number of words extracted from the document may be tens of thousands of number, which results in overhead. The purpose of the feature selection reduces the number of features by choosing the key terms and classifies the documents without performance degradation.

Chi-square is one of the most famous feature selection methods. The following is measurement of dependence of term t and category c.

$$\chi^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \tag{3}$$

Using the two ways contingency table of a term t and a category c, where A is the number of times t and c occur, B is the number of time the t occurs without c, C is the number of times c occurs without t, D is the number of times neither c nor t occurs, and N is the total number of documents. The χ^2 statistic has a natural value of 0 if c and t are independent [3].

3.2 Word2Vec

Word2Vec [4, 5] has enabled improvement of precision in the field of natural language processing as method for efficiently estimating the meaning of the word on a vector space. Word2Vec has begun from studying neural network and has premise that words with the same context have similar meaning [9]. Word2Vec represents each word in the vector space of dimension 200°. Word2Vec learns document and let each word to learn other words into neural network. Words of the associated meanings have high likelihood to appear nearby on the document and two words are getting closer vector by repeatedly learning.

4 Performance Evaluation

Data sets consist of 151 documents and its abstract is divided into papers which are relevant and not relevant to mental depression disease. We make the percentage of training data as 70 %, test data as 30 %. Then, we compare the accuracy of Chi-square with Word2vec. Before applying the Chi-square, value is given for the words by using HashingTF algorithm and then, the documents are classified by SVM and Naïve Bayes algorithm. When Word2vec is used, SVM and Naïve Bayes are applied immediately without HashingTF. Code of SVM and Naive Bayes is an API provided in Apache Spark. The Word2vec is compared with chi-square in Table 1. As shown in Table 1, when applying SVM and Naïve Bayes classification algorithms, we can see Word2vec

Table 1. Comparison Chi-square and Word2Vec

Accuracy (%) / Classification algorithms	Chi-square	Word2Vec
SVM	87.49	92.12
Naïve Bayes	88.23	N/A

is better than Chi-square in performance. Because Word2vec becomes vectors itself in multidimensional space where Word2vec has the meaning, the accuracy is high.

5 Conclusion

Chi-square is used to distinguish the vocabulary attributes of the particular topic. Since words belonging to Corpus have the characteristics of the corresponding corpus, the points are added or subtracted to the documents. After the calculation for all the words in the corpus, each word will have either one of the two characteristics to the contrary corpus and makes a list of words on the basis of this characteristic. Each word is stored in a list of words after words are assigned a tag for belonging to any corpus. Otherwise, Natural Language Processing (NLP) including Word2vec makes computers understand and analyze human language. Computers which could understand a meaning of word from a set of Unicode are able to distinguish the meaning of words itself with multi-dimensional vector method. That's why, the accuracy of Word2vec is higher than one of Chi-square in SVM and Naïve Bayes.

The input value of Naïve Bayes should be a positive value as any probability value. However, the output of Word2Vec includes the negative values, so we cannot apply the negative values to Naïve Bayes. Afterwards, we will try to find an appreciate classification algorithms to find out the performance of Word2Vec. Also, we used 151 documents, but to evaluate the performance of distributed process of Apache Spark, we plan to apply lots of documents.

References

1. Witten, I.H.: Text mining, Computer Science, University of Waikato, Hamilton, New Zealand, email: ihw@cs.waikato.ac.nz
2. Gupta, V., Lehal, G.S.: A survey of text mining techniques and applications. J. Emerg. Technol. Web Intell. **1**(1), 60–76 (2009)
3. Yang, Y., Pderson, J.O.: A comparative study on feature selection in text categorization. In: Proceedings of the 14th International Conference on Machine Learning (1997)
4. Mikolov, T., Chen, K., Corrado, G., Dean, J.: Efficient estimation of word representations in vector space (2013). arXiv preprint arXiv:1301.3781
5. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
6. Kotsiantis, S.B.: Supervised machine learning: a review of classification techniques. Informatica **31**, 249–268 (2007)
7. http://spark.apache.org/
8. Dean, J., Ghemawat, S.: MapReduce: simplified data processing on large clusters. Commun. ACM **51**(1), 107–113 (2008)
9. Le, Q.V., Mikolov, T.: Distributed representations of sentences and documents (2014). arXiv preprint arXiv:1405.4053

Analysis of Recent Maximal Frequent Pattern Mining Approaches

Gangin Lee and Unil Yun(✉)

Department Of Computer Engineering, Sejong University, Seoul, Korea
ganginlee@sju.ac.kr, yunei@sejong.ac.kr

Abstract. Since the concept of representative pattern mining was proposed to solve the limitations of traditional frequent pattern mining, a variety of relevant approaches have been developed. As one of the major techniques in representative pattern mining, maximal frequent pattern mining provides users with a smaller number of more meaningful pattern mining results. In this paper, we analyze characteristics of recent maximal frequent pattern mining methods using various concepts and techniques.

Keywords: Data mining · Knowledge discovery · Maximal frequent pattern · Pattern mining · Representative pattern

1 Introduction

The concept of representative pattern mining was proposed to overcome the fatal problems of previous traditional frequent pattern mining such as generating an excessive number of frequent patterns and degrading mining performance. Maximal frequent pattern mining [4, 11, 12], which is one of the major techniques in representative pattern mining, is known for extracting representative patterns more efficiently at the cost of accuracy in pattern restoration. Such an advantage has attracted development of various relevant applications such as bio data analysis [2, 9], uncertain data analysis [5], privacy preserving [6], distributed processing [7], social network analysis [10], and hypergraph dualization [8]. In this paper, we analyze characteristics of recent maximal frequent pattern mining techniques.

The remainder of this paper is as follows. Section 2 introduces the basic concept of frequent pattern mining and its important related works. Section 3 describes recent approaches of maximal frequent pattern mining. Section 4 finally concludes this paper.

2 Frequent Pattern Mining

Since the Apriori algorithm was devised [1], various frequent pattern mining approaches have been proposed. FP-Growth [3] is a tree-based approach that can solve the fatal problems of Apriori such as excessive database scans and candidate pattern creation. Its own tree structure, called FP-tree, and mining techniques have attracted many research attentions. As a result, a variety of variations and applications have been developed. From given databases, frequent pattern mining methods find all of the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_135

possible patterns such that their support values are not smaller than user-given minimum support threshold. Frequent patterns can be expressed as the following definitions.

Definition 1. (*Support of a pattern*) A database, $DB = \{T_1, T_2, \ldots, T_n\}$, is composed of multiple transactions, Ts, and each transaction $T_k = \{i_1, i_2, \ldots, i_m\}$ also has a number of items, is. Then, pattern $A = \{i_1, i_2, \ldots, i_j\}$, which can be generated from DB, is a subset of at least one T. The support of A, $Sup(A)$, is calculated as follows:

$$Support(A) = \sum_{i=1}^{n} f(A, T_i), f(A, T_i) = \begin{cases} 1, \text{ if } A \subseteq T \\ 0, \textit{ otherwise} \end{cases} \quad (1)$$

Definition 2. (*Frequent pattern*) A measure, called *minimum support threshold* (denoted as δ), is set by a user and employed to judge whether or not a pattern is frequent. This measure is the product of a user-given percent value and the number of transactions in DB. Then, pattern A is considered as a frequent pattern if the following condition is satisfied:

$$Support(A) \geq \delta \quad (2)$$

Therefore, the main goal of frequent pattern mining is to find all of the possible patterns that satisfy the above condition from a given database.

However, they may extract an excessive number of pattern results depending on features of databases and threshold settings. Since it is difficult to analyze all the mined patterns, we need to consider another approach.

3 Recent Maximal Frequent Pattern Mining Techniques

Maximal frequent pattern mining is a method that mines a smaller number of representative patterns instead of extracting all of the possible frequent patterns. There are two ways to mine representative patterns: closed frequent pattern mining and maximal frequent pattern mining. Although they can extract pattern mining results that can represent frequent patterns, maximal frequent pattern mining can guarantee better pattern condensing effect. Closed frequent pattern mining guarantees complete pattern restoration from closed frequent patterns to original frequent patterns; however, its pattern condensing effect is worse than that of maximal frequent pattern mining. This paper focuses on recent techniques of maximal frequent pattern mining. The maximality feature of a pattern is defined as follows.

Definition 3. (*Maximal frequent pattern*) Let X be a pattern, $\Gamma = \{X'_1, X'_2, \ldots, X'_k\}$ be a set of supersets of X, X's, and δ be a user-given minimum support threshold. Then, X becomes a maximal frequent pattern if the following conditions are satisfied:

$$Sup(X) \geq \delta, \Gamma = \{X' | Sup(X') < \delta\} \quad (3)$$

Through these constraints, we can effectively reduce the number of generated patterns.

The maximality characteristic has the following effect. The left side of Fig. 1 shows an example of frequent patterns. The right side of Fig. 1 is a result of expressing the frequent patterns as representative ones using the maximality feature of them. As shown in the figure, a large number of frequent patterns can be expressed as a smaller number of maximal frequent patterns according to their support characteristics.

$WFPmax_{WA}$ and $WFPmax_{SD}$ [12] are tree-based algorithms that extract maximal frequent patterns considering weight factors of items. They use a recursive divide-and-conquer manner and employ different item sorting orders according to algorithm types. They can effectively be utilized in various areas dealing with static data. Theses algorithms scan a given database twice in order to extract weighted maximal frequent patterns. Their basic frameworks follow that of FP-Growth [3]. Therefore, in the process of the first database scan, they calculate a weight ascending order (in the case of $WFPmax_{WA}$) or support descending order (in the case of $WFPmax_{SD}$) from the given data and prune invalid items. In the second database scanning process, they construct their own tree structures and perform pattern growth works recursively. IM_WMFI [11] is a weighted maximal frequent pattern mining algorithm for processing incremental databases. In contrast to the above ones, the method can be employed effectively in environments where data are continually accumulated. It also follows a tree-based pattern growth manner. Additionally, IM_WMFI constructs its own tree structure and performs mining operations within a single database scan in order to handle such dynamic data efficiently. Since all of the necessary works for mining weighted maximal frequent patterns have to be conducted within a single database scan, the algorithm directly stores given data into its own tree structure at first, and then it performs additional tasks for tree restructuring. In order to increase efficiency of tree restructuring, IM_WMFI divides its tree (the entire task) into a number of paths (smaller tasks) and performs item resorting operations for each path (a divide-and-conquer manner). AWMax [4] is an approach integrating the concepts of approximate pattern mining and maximal frequent pattern mining. The method mines maximal frequent patterns considering error tolerance and weight conditions on noise environments. Accumulated data may have unexpected errors such as noise, device

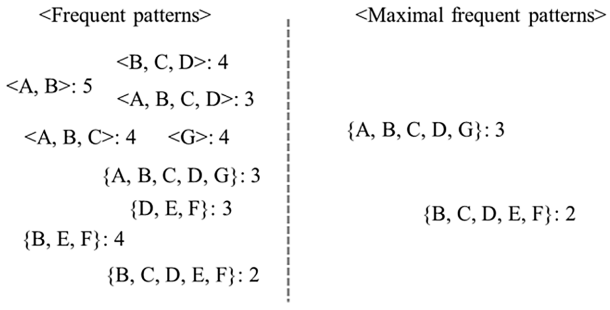


Fig. 1. Example of relations among frequent patterns and maximal frequent patterns

Table 1. Features of recent maximal frequent pattern mining algorithms

Algorithm	Stream processing	Considering weight factor	Tree-based approach	Approximate method
$WFPmax_{WA}$	N	Y	Y	N
$WFPmax_{SD}$	N	Y	Y	N
IM_WMFI	Y	Y	Y	N
AWMax	N	Y	Y	Y

malfunction or failure, record error, etc. In these cases, AWMax can effectively extract weighted maximal frequent patterns considering error tolerance.

Table 1 shows major features of the aforementioned algorithms. According to types of given data and purposes of pattern mining results, users can select and utilize proper algorithms.

4 Conclusion

In this paper, we explained the concept and characteristics of maximal frequent pattern mining and conducted analysis of recent relevant methods. Extensive usability of maximal frequent pattern mining led to various techniques and applications. Therefore, users can select and employ appropriate algorithms according to their own situations. The maximality property of patterns can also effectively be applied in various areas that may cause excessive computational overheads such as graph pattern mining, distributed processing of big data, etc. We are scheduled to conduct such expanded studies in our future work.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624) and the Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2015 (Grant no. C0261068).

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
2. Goparaju, A., Brazier, T., Salem, S.: Mining representative maximal dense cohesive subnetworks. Netw. Model. Anal. Health Inf. Bioinform. **4**(1), 29 (2015)
3. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Mining Knowl. Discov. **8**(1), 53–87 (2004)
4. Lee, G., Yun, U., Ryang, H., Kim, D.: Approximate maximal frequent pattern mining with weight conditions and error tolerance. Int. J. Pattern Recogn. Artif. Intell. **30**(6), 1650012:1–1650012:42 (2016)

5. Li, H., Zhang, N.: Probabilistic maximal frequent itemset mining over uncertain databases. In: 21st International Conference on Database Systems for Advanced Applications, pp. 149–163 (2016)
6. Karim, M., Rashid, M., Jeong, B., Choi, H.: Privacy preserving mining maximal frequent patterns in transactional databases. In: 17th International Conference on Database Systems for Advanced Applications, pp. 303–319 (2012)
7. Necir, H., Drias, H.: A distributed maximal frequent itemset mining with multi agents system on bitmap join indexes selection. Int. J. Inf. Technol. Manage. **14**(2/3), 201–214 (2015)
8. Nourine, L., Petit, J.: Extended dualization: application to maximal pattern mining. Theor. Comput. Sci. **618**, 107–121 (2016)
9. Salem, S., Ozcaglar, C.: MFMS: maximal frequent module set mining from multiple human gene expression data sets. In: 12th International Workshop on Data Mining in Bioinformatics, pp. 51–57 (2013)
10. Stattner, E., Collard, M.: MAX-FLMin: an approach for mining maximal frequent links and generating semantical structures from social networks. In: 23rd International Conference on Database and Expert Systems Applications, pp. 468–483 (2012)
11. Yun, U., Lee, G.: Incremental mining of weighted maximal frequent itemsets from dynamic databases. Expert Syst. Appl. **54**, 304–327 (2016)
12. Yun, U., Lee, G., Lee, K.: Efficient representative pattern mining based on weight and maximality conditions. Expert Syst. (2016). (in press)

Design of Noise Information Storage System Using IoT Devices

Judae Lee and Unil Yun(✉)

Department of Computer Engineering, Sejong University, Seoul, Korea
jdlee@sju.ac.kr, yunei@sejong.ac.kr

Abstract. Recently, many issues caused by noise have been on the rise in quite places such as reading room and a library where personal desks are used. In this paper, we design a system for collecting and storing noise-data into a database using Internet of Things (IoT). The data can be used for restraining the issues and addressing noise sources. Thus, our system performs sensing raw noise with IoT devices, transmitting noise-data from IoT devices to a server, processing the noise-data at the server, and storing the processed data in a database.

Keywords: IoT · Noise sensing · Noise information storage

1 Introduction

We design a noise information storage system to efficiently deal with the latest issues about noise from public places which have to be quite. Development of IoT makes it possible that things use wireless networks with low power. In our system, IoT devices are attached to desks or tables. The attached devices sense noise using noise sensors. After that, the noise data are processed and stored in a database.

2 Related Work

The internet of things (IoT) is the networks of physical devices—embedded with electronics, software, sensors, and networks connectivity that enable these objects to collect and exchange data. The Global Standards Initiative on Internet of Things (IoT-GSI) defined the IoT as "the infrastructure of the information society" [3].

2.1 Maintainability of IoT

The more IoT is used, the more maintainability of IoT gets important. Our proposed device has to get power from a battery and it is immobile. So, for the maintainability, consuming low power is an essential thing. There are, in addition, many important elements which have been researched [1, 4, 6, 9].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_136

2.2 Sound Source Localization

In our design, several devices can sense sound from one noise source. So, to point out one device which is the closest device with the noise source, noise source localization (SSL) is required. We can find direction of noise sources using SSL. SSL need more than two microphones to find direction of the noise sources. With two microphones, the direction of noise sources can be found on two-dimensional space. Over three microphones, three of those must not lie on a straight line, can be used for finding the direction of noise sources on three-dimensional space. Figure 1 shows abstract of SSL. Our purpose is to find noise sources from one floor and store it, so two-dimensional way will be used. Research for eliciting proper information using sound sensors have progressed [2, 5, 8].

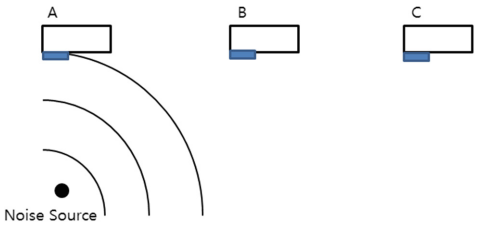


Fig. 1. Abstract of SSL

3 System Design

In this section, we describe the definitions of key terms. An IoT device is a device which can sense raw noise and can use wireless networks. A noise source is a location where the noise comes from. 'IoT device, A' means the device is named A. Analyzing raw noise yields noise data which are digital data. A server collects noise data from IoT devices and processes the data. Processed data are stored in a database.

Figure 2 shows the overall system structure and data flow. The two devices A and B perceive the noise which occurs from the noise source, and change the noise which is analog signal to the digital signal. If the digital data do not satisfy a base value, the data will be abandoned. Otherwise, it will be changed to decibel (dB) values. There are three ways to arrange the base value. One way is a server-base way that the server sets base values and transmits the value to devices. Another way is an initial-fix way that IoT devices have a initially fixed value before being used. The other way is environment-base that devices arrange a base value itself after being installed, based on average sound level from macro-environment of the devices.

The noise data include a dB value and their networks information which is employed for identifying neighbor devices. The server collects data from devices, waiting some period of time before processing the noise data. The period will have efficiency when the value is speculated based on relation between sound level and distance. More period of time than some delay which is required to be lower sound

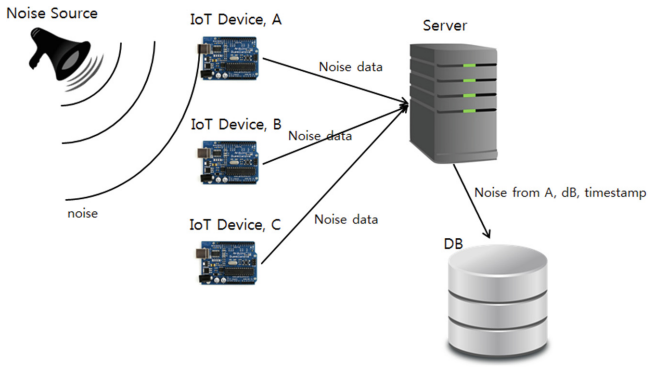


Fig. 2. A summary design of proposed system

level than the base value from the sound level of a noise source by means of distance will be useless. Figure 3 shows formula of sound attenuation and image concerned with it. With the formula, we can calculate the time when the sound level gets lower than the base value. If the best away distance of two devices in same area is not enough, sound propagation time between the two devices should be proper to be set as period. In addition, the different environment can lead to different period, so the server has to apply different period based on local environment when the server processes an identifying step. The period elicited from environment can be gathered by IoT devices or the period can be arranged before the first communication between IoT devices and the server by whom can manipulate the server. Too big period value can incur a problem that several noise occurrences are saved as one set of noise data.

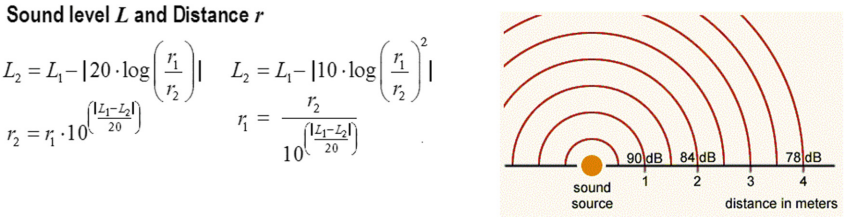


Fig. 3. Formulas for distance and sound level [7]

The identifying process includes analyzing of noise data to deduce the closest device to a noise source. In this procedure, the SSL will be used. After identifying the closest device, in this case A from Fig. 2, the noise data from A are stored in a database with timestamp and device's information. Figure 4 is a use case diagram of the system which contains the described process. The actor of IoT device is just supposed as desk. It can be other objects.

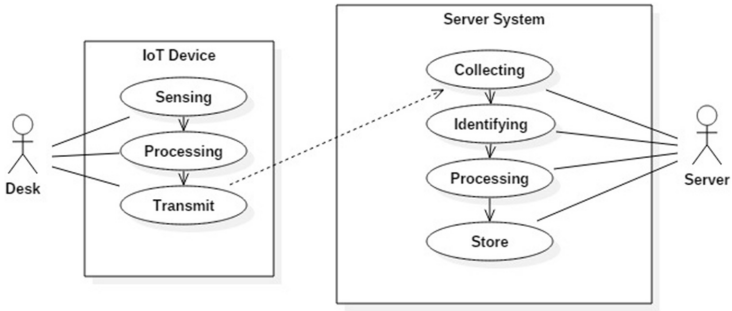


Fig. 4. Use case of proposed system

Figure 5, the algorithms bellow, shows the collection and identification parts of the use case. The collection procedure receives noise data from networks, and adds the noise data to a queue constantly. The identification procedure keeps surveillance on the queue, and identifies the closest IoT device from the noise source. After that, the selected data are processed in the processing step. Others which are estimated as far from the noise source are deleted from the queue.

```
Procedure collecting ( )
 Input: queue Q
 Output: data added queue Q
while(true)
    receive noise data ND from network
    NS.noisedata ← ND                          //NS is a structure
    NS.TimeStamp ← current time
    Q.enqueue ( NS )

Procedure identifying ( )
 Input: queue Q, some period of time p
 Output: selected data nd
while(true)
    td ← time defference between Q.front.TimeStamp and current time
    if (td > p)
        nd ← identified data from Q using SSL
        for c ← Q.front to Q.rear
            if ( nd.noisedata.network = c.noisedata.network)
                tdc ← time defference between nd and c
                if ( tdc < p )
                    Q.remove ( c )
        processing ( nd )
```

Fig. 5. Collecting & identifying method

4 Conclusion

In this paper we proposed a system which can gather noise data from specific spaces using IoT. By using this system, we can get the utilizable data. The data can be utilized in various systems, such as real-time noise monitoring system and current state of noise analyzing system.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624) and the Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2015 (Grant no. C0261068).

References

1. Chen, Y.-K., Wu, A.-Y.A., Bayoumi, M.A., Koushanfar, F.: Editorial low-power, intelligent, and secure solutions for realization of Internet of Things. IEEE J. Emerg. Sel. Top. Circuits Syst. **3**(1), 1–4 (2013)
2. Dehkordi, M.B., Abutalebi, H.R., Taban, M.R.: Sound source localization using compressive sensing-based feature extraction and spatial sparsity. Digital Signal Process. **23**, 1239–1246 (2013)
3. ITU. https://www.itu.int/en/ITU-T/studygroups/2013-2016/20/Pages/default.aspx
4. Khana, M., Dinb, S., Jabbar, S., Gohar, M., Ghayvat, H., Mukhopadhyaye, S.C.: Context-aware low power intelligent SmartHome based on the Internet of Things. Comput. Electr. Eng. **52**, 208–222 (2016)
5. Lam, J., Kapralos, B., Kanev, K., Collins, K., Hogue, A., Jenkin, M.R.M.: Sound localization on a horizontal surface: virtual and real sound source localization. Virtual Reality **19**(3–4), 213–222 (2015)
6. Porambage, P., Ylianttila, M., Schmitt, C., Kumar, P., Gurtov, A.V., Vasilakos, A.V.: The quest for privacy in the Internet of Things. IEEE Cloud Comput. **3**(2), 36–45 (2016)
7. USFS. http://www.fs.fed.us/t-d/programs/im/sound_measure/helo_index.shtml
8. Yook, D., Lee, T., Cho, Y.: Fast sound source localization using two-level search space clustering. IEEE Trans. Cybern. **46**(1), 20–26 (2016)
9. Zhao, C., Liu, J., Shen, F., Yi, Y.: Low power CMOS power amplifier design for RFID and the Internet of Things. Comput. Electr. Eng. **52**, 157–170 (2016)

Analysis of Privacy Preserving Approaches in High Utility Pattern Mining

Unil Yun(✉) and Donggyu Kim

Department of Computer Engineering, Sejong University, Seoul, Korea
yunei@sejong.ac.kr, donggyukim@sju.ac.kr

Abstract. With the significant increase of information sharing in various areas, it has been an important issue to prevent personal information from being disclosed to abnormal users. Pattern mining is one of data mining technique for extracting interesting pattern information from massive databases. Therefore, sensitive patterns belonging to personal information can be disclosed to abnormal users through pattern mining methods. A sanitization approach that modifies a given database is one of the most common approach for achieving privacy preserving. In this paper, we introduce and analyze various methods for achieving privacy preserving in high utility pattern mining based on sanitization approaches.

Keywords: Data mining · Utility pattern mining · High average-utility pattern mining · Stream pattern mining

1 Introduction

One of the data mining approaches, pattern mining, extracts meaningful pattern information from large-sized databases. In particular, high utility pattern mining [4, 5] can find more meaningful patterns compared to traditional frequent pattern mining, [3] because it considers the importance and quantity of each item (called utility information) in its mining process.

Various approaches for preserving sensitive patterns from being disclosed (called privacy preserving data mining (PPDM)) also have been proposed in order to achieve privacy preserving processes with minimum side effects and efficient performances. As shown in Fig. 1 [8], PPDM methods can be fallen into two categories, a data-sharing approach and a pattern-sharing approach. A pattern-sharing approach modifies the result pattern information in order to hide sensitive information contained in the result. On the other hand, a data-sharing approach modifies a database in order to prevent sensitive information from being mined by pattern mining methods. In a data-sharing approach, the most common approach is a sanitization approach, which manipulates the information of a database so that any sensitive pattern cannot be mined from the database.

Most of previous studies [6, 7] focus on hiding sensitive patterns from frequent pattern mining approaches [1, 3]. They manipulate the binary information of items in a given transaction database in order to make sensitive patterns have supports lower than a given support threshold. Furthermore, various methods for achieving PPDM in high utility pattern mining (called privacy preserving utility mining (PPUM)) have been

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_137

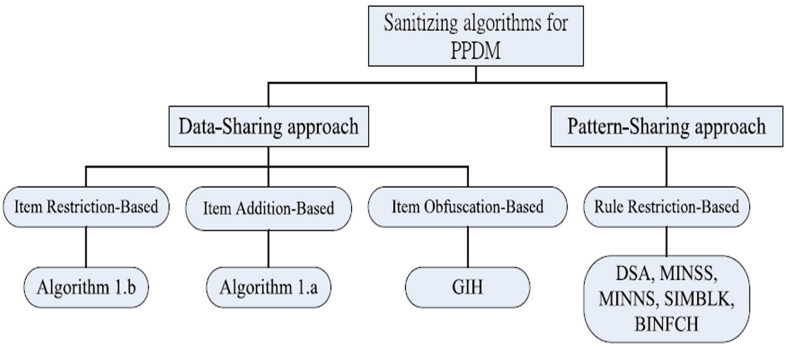


Fig. 1. PPDM approaches

developed in recent years. In this paper, we analyze state-of-the-art PPUM methods based on the sanitization approach.

The rest part of this paper is composed of the following contents. In Sect. 2, the explanation on related works such as high utility pattern mining is provided. In Sect. 3, we introduce and analyze state-of-the-art PPUM methods. Finally, in Sect. 4, we conclude the paper.

2 Related Works

High utility pattern mining calculates the utilities of patterns and extracts patterns with utilities no less than a given threshold called *minutil* as high utility patterns. The utility of a pattern is calculated as follows. Each item in a database has two elements, an internal utility and external utility.

An internal utility indicates the quantity of the item in a transaction. On the other hand, an external utility indicates the importance of item in the database. The utility of an item in a transaction can be calculated by the product of its internal utility and external utility. Then, the utility of a pattern can be obtained by summating all utilities of items contained in the pattern.

For example, consider a market database containing the information of customers. Each transaction presents the list of products purchased by the customer. Each product has its own unique prices. In addition, the quantity of each product in transactions can be different depending on customers.

As described in the above example, since high utility pattern mining approach can reflect the characteristics of the real world better than the traditional frequent pattern mining approach, numerous algorithms using an apriori approach and a pattern growth approach have been proposed in the past few decades.

3 Privacy Preserving Utility Mining Approaches

The overall procedure of the PPUM algorithm using a sanitization approach is shown in Fig. 2. First of all, a set of sensitive patterns determined based on privacy policy is inputted by users. In addition, an original database needed to be sanitized is also inputted. Then, PPUM processes for sanitizing the original database is conducted by manipulating the utility information of certain items. In these processes, algorithms generally reduce the internal utilities of items in order to decrease the utilities of sensitive patterns that containing the items. If all sensitive patterns have utilities less than *minutil* after such sanitization processes, a PPUM task is terminated and a modified database is obtained. Otherwise, sanitization processes are conducted repeatedly until all sensitive patterns have utilities less than *minutil*. As mentioned earlier, these processes are performed by changing the information of the original database. Therefore, various side effects such as the losses and generation of unintended patterns can be occurred during the sanitization processes. Due to such problem, the most important factor for evaluating the performances of PPUM methods is the degree of side effects caused during the sanitization processes.

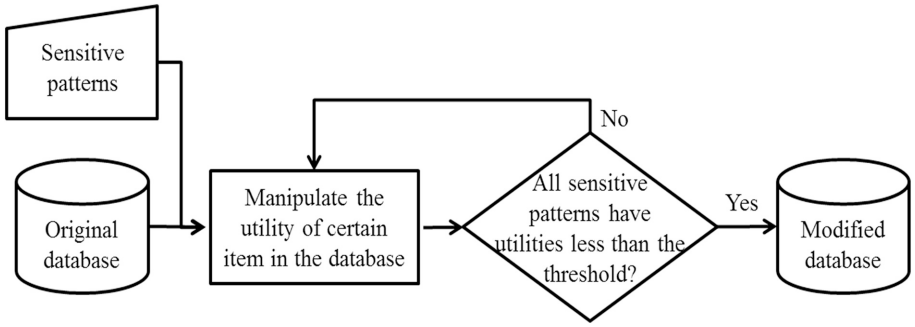


Fig. 2. Overall procedure of the sanitization approach

In [8], two algorithms for PPUM, HHUIF and MSICF, are proposed. They use different heuristic manners for manipulating a given database without causing significant side-effects. Their sanitization approaches are as follows. HHUIF firstly decreases the utility of an item having the highest utility among all items. On the other hand, MSICF firstly decreases the utility of an item contained in sensitive patterns more frequently than other items. Since above two algorithms perform their PPUM processes by scanning databases repeatedly, their runtime performances are not good.

FPUTT [9] is thus proposed to speed up the execution time of HHUIF and MSICF by using additional data structures. The algorithm scans a given database and constructs data structures called FPUTT-Tree, SI-table, and II-table in order to capture all information used for conducting sanitization processes without additional database scans. On the other hand, this algorithm adopts the heuristic manner of HHUIF in order to hide sensitive patterns in the database.

Even though above three algorithms employ heuristic manners in their sanitization processes so as to reduce side effects, the side effects caused by the algorithms are still considerable because they employ heuristic manners for manipulating databases and the sanitization process is a NP-Hard problem.

To solve this problem, PPUMGAT+ [2] is designed. PPUMGAT+ adopts a genetic algorithm, which imitates a human gene in order to find optimal solutions for problems. Therefore, PPUMGAT+ can find optimal ways for manipulating databases in such a way that side-effects are minimized. Table 1 shows the comparison of the algorithms.

Table 1. Comparison of PPUM algorithms

Algorithm	Database scanning	Sanitization approach	Use of data structure
HHUIF	Constantly	Highest utility	None
MSICF	Constantly	Highest conflict count	None
FPUTT	Two times	Highest utility	FPUTT-Tree, SI-table, II-table
PPUMGAT+	Constantly	Genetic algorithm based	None

4 Conclusion

In this paper, we introduced and analyzed state-of-the-art PPUM approaches. Various heuristic methods have been proposed in order to hide sensitive patterns by minimizing side effects. In addition, most recent methods adopt a genetic algorithm in order to solve the NP-hard problem of a sanitization process. Since these algorithms still cause considerable side-effects on original databases, more researches on reliable PPUM methods should be conducted in the future.

Acknowledgments. This research was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF No. 20152062051 and NRF No. 20155054624) and the Business for Academic-industrial Cooperative establishments funded Korea Small and Medium Business Administration in 2015 (Grant no. C0261068).

References

1. Agrawal, R., Srikant, R.: Fast algorithms for mining association rules. In: 20th International Conference on Very Large Data Bases, pp. 487–499 (1994)
2. Lin, J.C.-W., Gan, W., Fournier-Viger, P., Yang, L., Liu, Q., Frnda, J., Sevcik, L., Voznak, M.: High utility-itemset mining and privacy preserving utility mining. Perspect. Sci. **7**, 74–80 (2016)
3. Han, J., Pei, J., Yin, Y., Mao, R.: Mining frequent patterns without candidate generation: a frequent-pattern tree approach. Data Min. Knowl. Disc. **8**(1), 53–87 (2004)
4. Kim, D., Yun, U.: Efficient mining of high utility pattern with considering of rarity and length. Appl. Intell. **45**(1), 152–173 (2016)
5. Ryang, H., Yun, U.: High utility pattern mining over data streams with sliding window technique. Expert Syst. Appl. **57**(15), 214–231 (2016)

6. Verykios, V.S., Elmagarmid, A.K., Bertino, E., Saygin, Y., Dasseni, E.: Association rule hiding. IEEE Trans. Knowl. Data Eng. **16**(4), 434–447 (2004)
7. Wu, Y.-H., Chiang, C.-M., Chen, L.P.: Hiding sensitive association rules with limited side effects. IEEE Trans. Knowl. Data Eng. **19**(1), 29–42 (2007)
8. Yeh, J.-S., Hsu, P.-C.: HHUIF and MSICF: novel algorithms for privacy preserving utility mining. Expert Syst. Appl. **37**, 4779–4786 (2010)
9. Yun, U., Kim, J.: A fast perturbation algorithm using tree structure for privacy preserving utility mining. Expert Syst. Appl. **42**, 1149–1165 (2015)

An Agent-Based Remote Operation and Safety Monitoring System for Marine Elevators

Hyung-Joo Kim(✉) and Kwangil Lee

ICT Convergence Security Research Team, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon 34129, Korea
{kimhj, leeki}@etri.re.kr

Abstract. The reliability of marine elevators are important because of the tough conditions in the marine environment. For the safety purpose, more than 26 safety lists shall be checked by persons when the ship is docked. However, its maintenance costs of the elevator increases as the rise of personnel expenses increases. Also the efficiency is highly dependent on the ability of the personnel but it is not feasible to provide expert maintenance services all the time. Therefore we propose an agent-based remote operation and safety monitoring system for the marine elevators through a shipborne gateway, called MariComm. So expert A/S personnel can inspect status of marine elevators and provide best maintenance service through the marine elevator management server whenever and wherever he wants. In our experimental analysis, we illustrated that the performance and reliability of the elevator can be improved through analyzing the accumulated data.

Keywords: Marine elevator · Remote monitoring · Safety monitoring · Elevator agent

1 Introduction

The marine elevator has been required real-time safety inspections periodically so as to operate safely and reliably in poor marine environments. So far, its work belongs to the expert personnel and the ability of the personnel determines inspection work efficiency. The general maintenance routine of the marine elevators proceeds as follows. First, the ship needs to be docked at the port. Then the management team waiting a ship goes aboard it and checks status of it. And if necessary, the A/S team goes aboard a ship and repairs failure of it. So it takes quite long time since it is too complex and inefficient.

Recently, the communication environment at the shipborne is rapidly improved in the throughput and coverage as maritime communication technologies evolves [1, 2, 3]. ETRI developed MariComm (Maritime broadband Communication) system that is able to provide broadband internet/multimedia services available at a rate of 1 Mbps or more on sea [4]. Therefore, it is possible to have an IT-based maintenance and monitoring system remotely to systematically manage component lifecycle, fault diagnosis, safety operations, and status data of elevators. For this, we propose and implement an agent-based remote operation and safety monitoring system in this paper.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_138

2 Requirements

We investigate on the user and functional requirements for a remote monitoring and maintenance system for the marine elevators.

2.1 User Function Requirements

The user function requirements for remote operation and safety monitoring are as follows.

- Monitoring, saving, and visualizing the status of the marine elevators
- Checking the status of the marine elevators by an elevator inspector
- Remotely checking the status of the marine elevators by service users (manufacturer, ship owner, ship builder, maintenance co.)
- Having timestamp and elevator ID in the collected data

2.2 Operation and Safety Monitoring Items

The main operation and safety monitoring items are as follows.

- Current floor, running status, and operation mode of an elevator
- Position of an elevator with GPS
- Measurement of safety gap between floor door and elevator shaft with inductive proximity sensors(15 mm)
- Measurement balancing of elevator cage with gyro enhanced digital compass
- Measurement load weight with multi-channel shear beam load cells(500 kg)
- Detecting door open/close status with door interlock switch
- Checking the tightening of wire rope clip with magnetic sensor
- Measurement environmental conditions with sensors
- Detecting fire with UV/IR composite flame detector

3 System Architecture

3.1 The Concept of the Remote Operation and Safety Monitoring

The remote operation and safety monitoring system consists of MariComm, marine elevator agent, and marine elevator management server. MariComm provides broadband communication services with the data rate of 1 Mbps or more in sea by applying multi-hop relay technique to extend the range of the wireless communication. The marine elevator agent gives functions of message interconnection with an elevator controller, saving data of operation and safety, transferring data to the marine elevator management server, and displaying data of operation and safety on a monitor. The marine elevator management server has functions of saving data of operation and safety, fault diagnosis of main safety components, and web service for managing

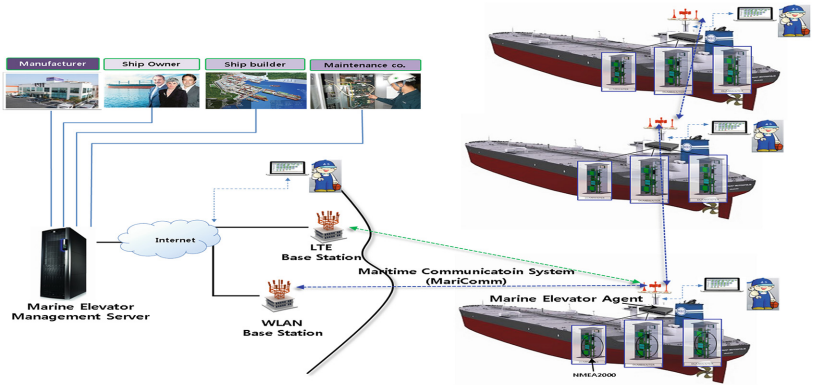


Fig. 1. The remote operation and safety monitoring for marine elevators

elevator safety remotely. So the elevator A/S personnel can inspect status of marine elevators through the marine elevator management server whenever and wherever he wants. Therefore he can provide a fast and expert maintenance service of marine elevators. Then improved performance and reliability of the elevator can be provided through analyzing the accumulated data (Fig. 1).

3.2 The System Diagram and SW Components of the Marine Elevator Agent

The system diagram of the marine elevator agent is shown in Fig. 2. It consists of data collection module, data storage module, transmission module and visualization module. The data collection module collects nmea0183 message-typed data through ethernet communication from the elevator controller. The data storage module saves data in internal DB after data parsing process. The visualization module displays status information of operation and safety through a monitor, local PCs, and portable terminals. The transmission module transfers data to the marine elevator management server through MariComm system (LTE or Wi-Fi).

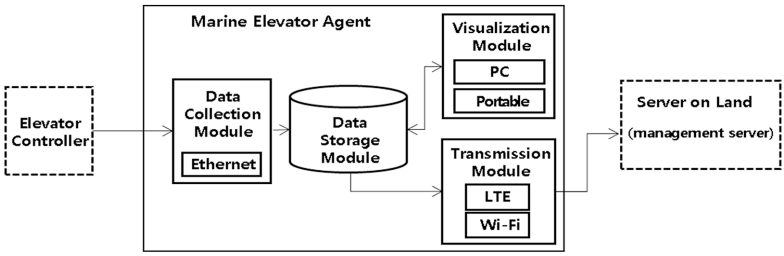


Fig. 2. The system diagram of the marine elevator agent

The SW components of the marine elevator agent are shown in Table 1.

Table 1. SW components and role of the marine elevator agent

Components	Role
Agent	– Save the collected data from main controller in DB – Transfer real-time data to the marine elevator management server – Transfer the stored data to the marine elevator management server – Transfer the status of communication with main controller to the marine elevator management server
DB	– Keep the collected data for a predetermined time – Save data to display from the web server
Web server	– Serve web service to check data of operation, safety, GPS, temperature/humidity/barometer, IMU and load cell through web browser remotely
GUI	– Express data of operation, safety, GPS, temperature/humidity/barometer, IMU and load cell

3.3 The System Configuration for the Agent-Based Remote Operation and Safety Monitoring

The system configuration for the agent-based remote operation and safety monitoring is shown in Fig. 3. Elevator controller consists of main controller, car controller, and floor controller. Main controller collects GPS data and check the tightening of wire rope clip. It also aggregates data of car controller and floor controller through NMEA2000 network. Then it transfers them to the marine elevator agent through ethernet network. Car controller manages data of fire sensor, temp/humidity/barometer sensor, IMU, load cell, limit switch, proximity sensor, and door interlock switch. It transfers data to main controller through NMEA2000 network.

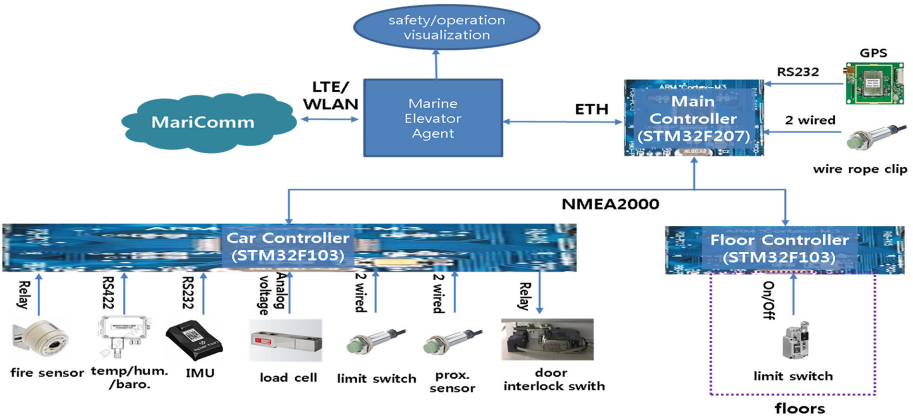


Fig. 3. The system configuration for the agent-based remote operation and safety monitoring

Floor controller checks the current floor of elevator cage with limit switch and transfers data to main controller through NMEA2000 network.

4 Experimental Results

4.1 Test Environment

We implemented the agent-based remote operation and safety monitoring system and established a test-bed as illustrated in Fig. 4. As shown in the figure, we installed a dumbwaiter elevator at the elevator test tower equipped with two floors. A marine elevator agent and a main controller are together installed at the elevator control panel. The LAM (Land Access Module) of MariComm was set up as station function of ship and SAM (Seaward Access Module) of MariComm as an access point function of shore. The two modules are connected to the internet through the gateway. Finally, we installed some monitoring sensors and devices in the elevator shaft and cargo.

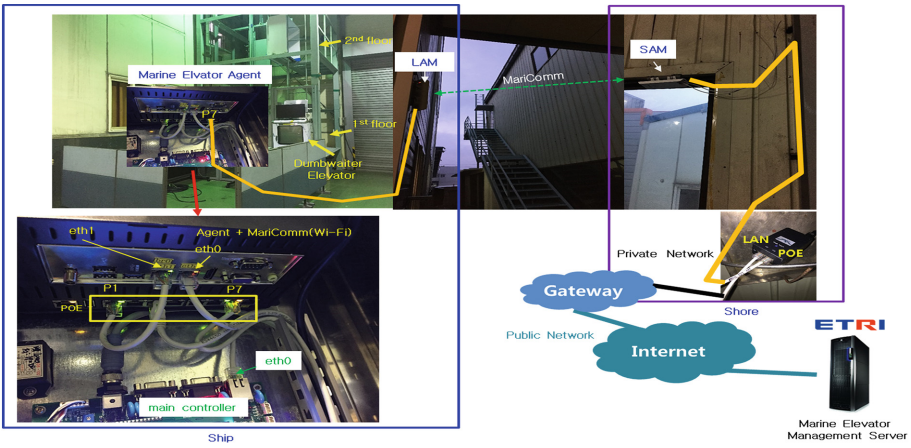


Fig. 4. Test environment for the agent-based remote operation and safety monitoring for the marine elevator

4.2 Test Results

We performed the experimental analysis for the functions of the agent-based remote operation and safety monitoring system while continuously testing the elevator's run/stop operations and the door's open/close operations. The main controller transferred the sensor data every 1 s including GPS information. Then the agent system refreshed the monitoring information every 1 s and transferred it to the marine elevator management server in a real time as it is shown in Fig. 5. And when the internet disconnected, the agent system stored data in internal DB. The stored data were transferred to the marine elevator management server when the internet is reconnected.

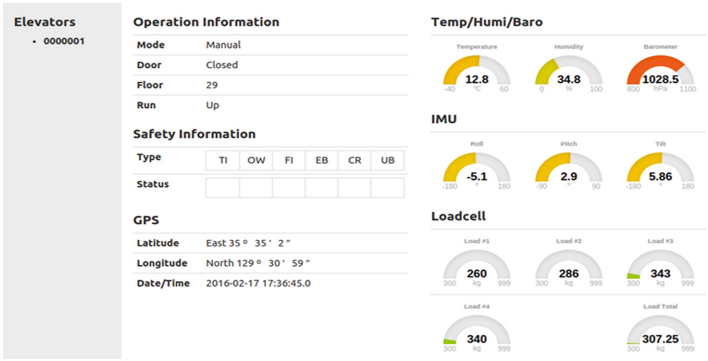


Fig. 5. Real-time data display of the marine elevator agent

5 Conclusion

This paper proposes an agent-based remote operation and safety monitoring for the marine elevators. The agent-based remote operation and safety monitoring consists of MariComm, marine elevator agent, and marine elevator management server. The marine elevator agent supports remote operation and safety monitoring of marine elevators through MariComm. Therefore, the elevator A/S personnel can provide a real-time expert maintenance service of marine elevators. Then enhanced performance and reliability of the elevator can be provided through analysis of the accumulated data. The intelligent maintenance service with machine learning is left for the future work.

Acknowledgments. The contents of this paper are the result of the Standard Technology Improvement Program of the Ministry of Trade, Industry and Energy of Korea (International Standard Development of Shipborne Common Data Model for Maritime IoT, 10058948).

References

1. ITU-R M. 1842-1. Characteristics of VHF radio systems and equipment for the exchange of data and electronic mail in the maritime mobile service RR Appendix 18 channels (2009)
2. IMO(International Maritime Organization). http://www.imo.org
3. Lee, K.I., Song, M.S., Jang, B.T.: International Standards and Technology for E-navigation and Maritime IoT. Electronics and Telecommunications Trends (2014)
4. Kim, H.-J., Choi, J.-K., Yoo, D.-S., Jang, B.-T., Chong, K.-T.: Implementation of MariComm bridge for LTE-WLAN maritime heterogeneous relay network. In: ICACT 2015, pp. 230–234 (2015)

Survey on CPN Applications in Cloud Computing

Rustam Rakhimov Igorevich(✉) and Dugki Min

DMSLAB, Computer Science Department, Konkuk University,
Hwayangdong, Gwangjingu, Seoul, Republic of Korea
{rustam, dkmin}@konkuk.ac.kr

Abstract. The multiple issues in a Cloud-computing environment require formal verification and validation methods. Dozens of research works have been done in order to refer this topic. In this paper we would like to emphasis the CPN (Colored Petri Nets) based formal analysis of different aspects in cloud computing. Research works regarding to CPN applications in cloud computing were categorized into four classes. The paper also covers the CPN modelling approaches specifically oriented to solve issues in cloud environments.

Keywords: CPN (Colored Petri Nets) · Cloud · Data-integrity · Model

1 Introduction

Considering dozens of existing works related to CPN we couldn't finalize the importance of it in research. We find there is no work that analyze and survey importance of CPN, especially from perspective of cloud related research works. In this work we try to collect and survey list of papers, where the CPN has been applied to the cloud environments. The conclusion has been astonishing, because CPN appears to be very useful and important tool that can be applied almost every sector of Cloud Computing research. We have classified the research papers where CPN is applied to the Cloud computing. As result we have derived four classes: Work-flow, Security, Service Composition and Analytical research works.

The paper structure goes as follows, in Sect. 2 we introduce shortly Colored Petri Nets and give useful references for further studies. Section 3 contains the main body of the paper where the main classification table is given. Some are the important approaches and references have been described in details. Finally the paper ends with conclusion about importance of CPN in Cloud computing environment.

2 Colored Petri Nets

The Petri Nets are popular and well-known formalism for modeling concurrency systems. Petri Net is the collection of basic elements such as places, transitions, arcs and tokens. Tokens occupy places and move to another place through arcs when corresponding transitions enabled.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_139

Colored Petri Nets (CPN) is an extended from original Petri Net and represents a well-known formalism for modeling concurrent protocols [20, 21]. CPN is applied in many areas where the concurrent and complex processes must be analyzed from architecture checking and behavior perspectives.

Colored Petri Nets (CP-nets or CPNs) [23–26] is a graphical language for constructing models of concurrent systems and analyzing their properties. CPN actively involved with ML programming language, which is based on the functional programming language Standard ML [27, 28], provides the primitives for the definition of data types, for describing data manipulation, and for creating compact and parameterizable models. CPN modeling language is developed as a general-purpose modeling language, aimed towards a very broad class of systems that can be characterized as concurrent systems. CP-nets mostly cover domains such as communication protocols, data networks, distributed algorithms, embedded systems and many other domains [29].

3 Importance of CPN in Cloud Computing

In order to answer the question how important the CPN for cloud computing research we have surveyed dozens of papers. First of all rough search with CPN and Cloud word combinations were performed. From the result list most of non-relevant research works were filtered out. Remaining research work list has been studied in detail in order to find some classification feature. Finally the classification has been found. To the best of our knowledge we have listed out most popular research works regarding to CPN application into cloud computing area. Table 1 is the final list of our survey that contains class name and corresponding papers.

Table 1. Categorization of cloud related research works utilized CPN

Class	Reference papers
Workflow	[10–13]
Security	[14–18]
Service Composition and Optimization	[19–24]
Analytical	[25–30]

Authors of [20] used CPN tools very efficiently and demonstrated the optimization of server consolidation for heterogeneous computer clusters. Server consolidation is an advanced topic in virtualization of cluster computing systems. Server consolidation has been and continuing to be a central research topic by many researchers. Dynamic server consolidation techniques engage with live migration of virtual machines (VMs) across the physical machines (PMs), in order to improve the manageability of clusters [20]. Main purpose of server consolidation is to reducing the cost of power, cooling and, hence contributes to grow of green data centers.

Dynamic flow sensitive security model for a federated cloud system has been introduced in [14]. The petri nets and the associated verification techniques were used to analyze the security of information flow. Specifically saying the CPN has been

applied to the part of Access Control sub-net, representing the interactions of the subjects and objects residing on the same cloud.

Basic structure of the access control sub-net model illustrated in Fig. 1 represents the data destruction and creation can be represented by simplified versions of the read and write transitions. Tokens represent entities, places represent different clouds and transitions capture the activities and security rules. s, l, etc., are parameters and f, f′, g, h are application specific functions. Those functions used to capture the ongoing computations and the resulting changes to data values and services. The model illustrated in Fig. 1 can be considered as a building block for designing complex chain of write and read access controls.

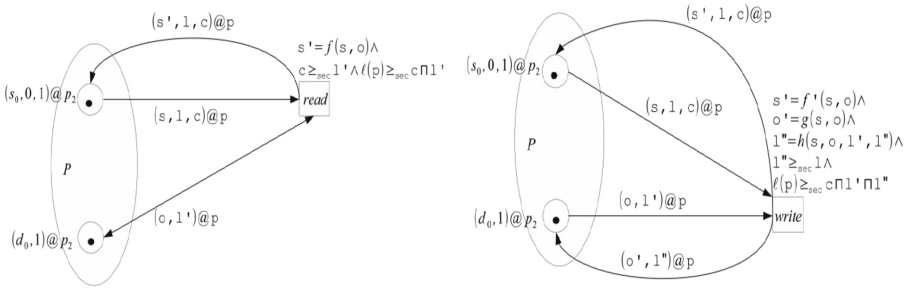


Fig. 1. Basic structure of the access control sub-net model

Novel secure mechanism for secure and fault-tolerant cloud-based information storage has been proposed in [15]. Security mechanisms in CPN model covers cloud service providers, service directories, service queries, key management and cluster restoration utilities. Compared to other works the authors of [15] designed quite simple model of the cloud storage. It conveys straightforward high-level CPN model of client, manager and cloud interaction.

Primitive high-level CPN model of the cloud storage illustrated in Fig. 2, where patient and doctor can be considered as a client blocks with different level of privileges and roles. Directory component is the instance of manager block that is responsible for administering the workflow. Details internal model of the directory component can be referred at [15].

The authors [24] have introduced context enabled CPN approach for validating a task migration in a pervasive cloud environment. This model does not focus on machine level, but more concentrates on tasks. In order to start migration process the CPU and RAM loads in mobile centers are watched. Physical migration of the tasks performed using OSGi bundles. Their system has been built on top of OSGi-based pervasive cloud [31, 32].

Another work related to data integrity in PaaS level of cloud environment. The authors [22] choose to apply CPN based approach to modeling and evaluating the data integrity using generalized transaction systems including cloud environments. Constraints on database records expressed in the form of predicate logic formulae and examined by CPN/ML codes implemented as a guard functions on CPN transitions.

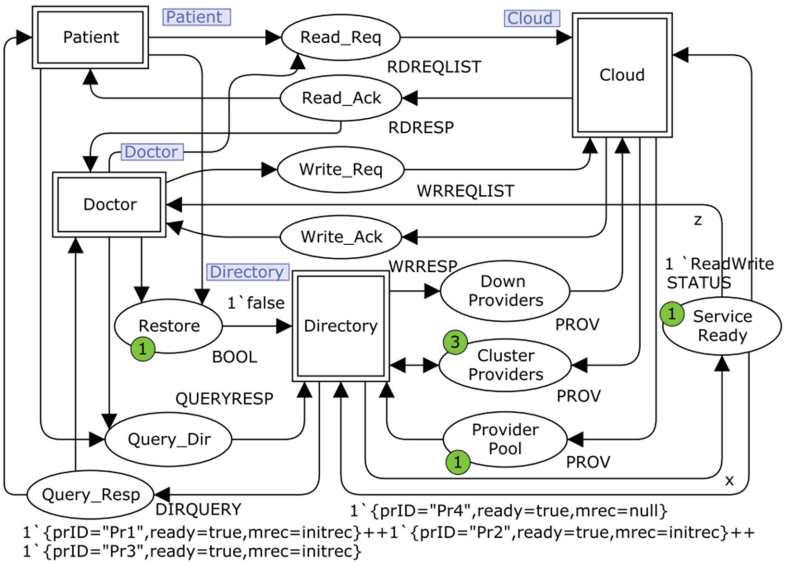


Fig. 2. High-level CPN model of the cloud storage

The CPN tool is not equipped with the verification capability, the constraint evaluation mechanism should be added to the model. The mechanism should be triggered at every "commit" or "abort" transactions, which makes the database update confirmed. Simple implementation of this mechanism is illustrated in Fig. 3. Depend on the "commit" or "abort" transitions the place C will have corresponding token. If the guard functions assuring the data integrity logically satisfied, then transition "A" will trigger and success message "OK" is forwarded to place "M". If the transition "D" with guards for detecting the breaking of data integrity is triggered, then error messages forwarded to place "N".

Evaluating data integrity in BASE transaction systems the simulation based and model driven approach was used by authors of [18]. CPN has been used to express and simulate the behavior of BASE transaction systems. Modeling data integrity has been performed in three orthogonal viewpoints. The first is the infrastructure viewpoint,

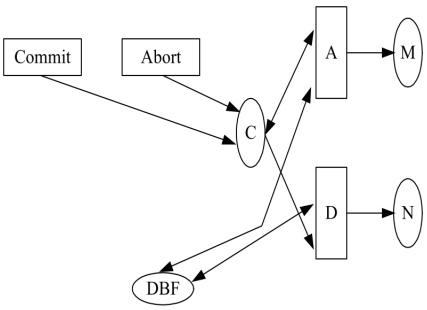


Fig. 3. Integrity evaluation mechanism

which represents the transaction processing mechanism including scheduling, queueing, concurrency control, database accesses and recovery control. Second viewpoint contains the structure of database accessed by each transaction. Finally third viewpoint covers the application logic in each transaction.

4 Conclusion

As a conclusion to our survey work we can say that modeling of the processes and work-flows can be performed in a straightforward or abstract level. When the model designed by researcher has been done in a straightforward way it does not expand well. In order to reach high level expansion of our model we need to think and design it in a more abstract level. It should be easily expandable and the number of participating nodes should not be drawn as a place. It should be designed as a token. If the basic Petri Nets are used then designing the model where single token represents the virtual or physical machine is not primitive. The CPN is the best solution to design models for Cloud computing problems, because single token can contain complex structure that can hold many important parameters of the machine.

Acknowledgement. This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the University Information Technology Research Center support program (IITP-2016-R2720-16-0004) supervised by the IITP(Institute for Information & communications Technology Promotion)

References

1. Petri, C.: Kommunikation mit Automaten. Bonn: Institut fur Instrumentelle Mathematik. Schriften des IIM, no. 2 (1962)
2. Reisig, W.: Petri Nets: An Introduction. EATCS Monographs on Theoretical Computer Science, vol. 4 (1985)
3. Jensen, K.: A brief introduction to Colored Petri Nets. In: Workshop on the Applicability of Formal Models, Aarhus, Denmark (1998)
4. Jensen, K.: An introduction to the theoretical aspects of Colored Petri Nets. In: Workshop on the Applicability of Formal Models, Aarhus, Denmark (1998)
5. Jensen, K.: Coloured Petri Nets; Basic Concepts, Analysis Methods and Practical Use. Basic Concepts, Monographs in Theoretical Computer Science, vol. 1 (1992)
6. Jensen, K.: Basic Concepts, Monographs in Theoretical Computer Science, vol. 1. Springer (1992). Analysis Methods, Monographs in Theoretical Computer Science, vol. 2 (1994)
7. Jensen, K.: Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use. Practical use, Monographs in Theoretical Computer Science, vol. 3 (1997)
8. Milner, R., Tofte, M., Macqueen, D.: The Definition of Standard ML. MIT Press, Cambridge (1997)
9. Ullman, J.D.: Elements of ML Programming. Prentice Hall, Englewood Cliffs (1998)
10. Bendoukha, S., Wagner, T.: Cloud transition: integrating cloud calls into workflow Petri Nets. In: CEUR Workshop Proceedings, vol. 851, pp. 215–216 (2012)
11. Li, Y., Boucelma, O.: A CPN provenance model of workflow: towards diagnosis in the cloud. In: CEUR Workshop Proceedings, vol. 789, pp. 55–64 (2011)

12. Hou, Q., Xie, Q., Li, S.: Development and application of lightweight cloud platform workflow system. Comput. Model. New Technol. **18**, 197–201 (2014)
13. Joo, K., Kim, S., Kim, D.: Cost-aware workflow scheduling scheme based on colored Petri-net model in cloud. In: International Conference on Future Web (ICFW 2014), Nov. 2014
14. Zeng, W., Kounty, M., Watson, P.: Formal verification of secure information flow in cloud computing. J. Inf. Secur. Appl. **27–28**, 103–116 (2016)
15. Fitch, D., Xu, H.: A Petri net model for secure and fault-tolerant cloud-based information storage. In: Seke, pp. 333–339 (2012)
16. Jasiul, B., Szpyrka, M., Sliwa, J.: Detection and modeling of cyber attacks with Petri nets. Entropy **16**, 6602–6623 (2014)
17. Huang, H., Kirchner, H.: Component-based security policy design with colored Petri nets. In: Palsberg, J. (ed.). LNCS, vol. 5700, pp. 21–42Springer, Heidelberg (2009). doi:10.1007/978-3-642-04164-8_3
18. Nishida, S., Shinkawa, Y.: Data integrity in cloud transactions. In: CLOSER 2014 – Proceedings of the 4th International Conference on Cloud Computing and Services Science, pp. 457–462 (2014)
19. Gutierrez-Garcia, J., Sim, K.: Agent-based service composition in cloud computing. In: Grid and Distributed Computing, Control and Automation: International Conferences, GDC and CA, pp. 1–10 (2010)
20. Al-Azzoni, I.: Server consolidation for heterogeneous computer clusters using Colored Petri Nets and CPN Tools. J. King Saud Univ. Comput. Inf. Sci. **27**(4), 376–385 (2015)
21. Ribas, M., Furtado, C., de Souza, J., et al.: A Petri net-based decision-making framework for assessing cloud services adoption: the use of spot instances for cost reduction. J. Netw. Comput. Appl. **57**, 102–118 (2015)
22. Shinkawa, Y.: CPN based data integirty evaluation for cloud transactions. In: Proceedings of the 7th International Conference on Software Paradigm Trends, pp. 267–272 (2012)
23. Suminto, R., Laaksono, A., Satria, A., et al.: Towards pre-deployment detection of performance failures in cloud distributed systems. In: Proceedings of the 7th USENIX Conference on Hot Topics in Cloud Computing (2015)
24. Zhu, L., Tan, S., Zhang, W., et al.: Validation of pervasive cloud task migration with colored Petri Net. Tsinghua Sci. Technol. **21**(1), 89–101 (2016). ISSN 1007-0214
25. Li, J., Cui, Y., Ma, Y.: Modeling message queueing services with reliability guarantee in cloud computing environment using Colored Petri Nets. Math. Prob. Eng. **2015**, 20 (2015) Hindawi Publishing Coroporation
26. Benmerzoug, D.: Designing complex agent interaction protocols using Colored Petri Nets: the cloud services composition case study. Intl. J. Agent Technol. Syst. **6**(2), 51–72 (2014)
27. Malik, S., Khan, S., Srinivasan, S.: Modeling and analysis of state-of-the-art VM-based cloud management platforms. IEEE Trans. Cloud Comput. **1**(1), 50–63 (2013)
28. Wang, M., Liu, Y.: QoS evaluation of cloud service architecture based on ANP. In: Proceedings of the International Symposium on the Analytic Hierarchy Process (2013)
29. Wang, Y., Chen, R., Wang, D.: A survey of mobile cloud computing applications: perspectives and challenges. Wireless Pers. Commun. **80**, 1–29 (2014)
30. Yim, J.: Design of a time Colored Petri Net model of the cloud-based mobile TV system. Intl. J. Softw. Eng. Appl. **9**(8), 241–252 (2015)
31. Shiraz, M., Gani, A.: A lightweight active service migration framework for computational offloading in mobile cloud computing. J. Supercomput. **68**(2), 978–995 (2014)
32. Zhang, W.S., Chen, L.C., Liu, X., Lu, Q.H., Zhang, P.Y., Su, Y.: An OSGi based adaptive and flexible pervasive cloud infrastructure. Sci. China Inf. Sci. **53**(3), 1–11 (2014)

An Observation Method for Estimating Carrier Frequency Offset in OFDM Systems

Mustafa Altaha(✉) and Humor Hwang(✉)

Myongji University, Yongin, Korea
hmhwang@mju.ac.kr

Abstract. Orthogonal Frequency Division Multiplexing (OFDM) is widely considered as an effective approach for current and future high-speed wireless communication. However, one of the main drawbacks of the OFDM is that it is sensitive to carrier frequency offset (CFO). The CFO causes interference among the multiplicity of carriers in the OFDM signal. Thus, the CFO must be estimated and compensated for in OFDM communication systems to minimize adverse effects of inter-carrier interference on the signal and maintain orthogonality. We propose a method based on observation training symbols for estimating CFO by employing block-by-block estimation. An analytical expression for the mean squared error of the frequency synchronization scheme is given and the results show that the proposed method has a superior performance compared to the conventional methods.

Keywords: Orthogonal Frequency Division Multiplexing (OFDM) · Carrier frequency offset (CFO) estimation

1 Introduction

One of the principal advantages of OFDM is its robust against multi-path channel which can cause inter-symbol interference (ISI) and inter-carrier interference (ICI). However, this is prevented in OFDM by the insertion of a cyclic prefix between successive OFDM symbols. Cyclic prefix (CP) or cyclic extension was first introduced by Peled and Ruiz in 1980 [1] for OFDM systems. In their scheme, conventional null guard interval is substituted by cyclic extension for fully-loaded OFDM modulation. As a result, the orthogonality among the subcarriers was guaranteed. With the trade-off of the transmitting energy efficiency, this new scheme can result in a phenomenal ISI reduction. Hence it has been adopted by the current IEEE standards. In 1980, Hirosaki introduced an equalization algorithm to suppress both ISI and ICI [2].

However, one of the main disadvantages of OFDM systems is sensitivity against carrier frequency offset (CFO), which causes inter-carrier interference. The sensitivity of the OFDM to the CFO in single carrier systems is a critical issue [3]. In general, the CFO is defined as the difference between the nominal frequency and actual output frequency. In OFDM, the uncertainty in carrier frequency due to a difference in the frequencies of the local oscillators in the transmitter and receiver gives rise to a shift in the frequency domain. This shift is also referred as frequency offset. It can also be caused due to Doppler shift in the channel. The demodulation of a signal with offset in

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_140

the carrier frequency can cause large bit error rate and may degrade the performance of a symbol synchronizer. Therefore, it is important to estimate the frequency offset and minimize or eliminate its impact [4].

Several techniques were proposed to estimate the carrier frequency offset in time domain and frequency domain. Moose [5] proposed a technique for maximum likelihood estimate (MLE) of frequency offset using the discrete fourier transform (DFT) values of a repeated data symbol. In this technique, the accuracy required of frequency offset correction depends on how much residual offset can be tolerated. The acquisition range of the technique is $\pm 1/2$ the intercarrier spacing of the repeated symbol. Schmidl and Cox [6] pro-posed frequency and timing synchronization algorithm by using repeated data symbols. The range of CFO estimation is ± 1. Beek, Sandell and Borjesson [7] proposed the joint maximum likelihood estimator of time and frequency offset in ODFM systems. The estimation uses the redundant information contained within the cyclic prefix. Zhou et al. [8] presented two maximum likelihood CFO estimation schemes, one in frequency domain and another in time domain, both under Doppler fading. Classen and Meyr [9] introduced a method to find both the symbol timing and carrier frequency offset. Pilot tones can be inserted in the frequency domain and transmitted in every OFDM symbol for the carrier frequency offset tracking. After estimating carrier frequency offset from pilot tones in the frequency domain, the signal is compensated with the estimated carrier frequency offset in the time domain.

In this paper, we propose a method based on observation training symbols S. These are grouped into two consecutive blocks, where each block has a length of $S/2$ and block-by-block estimation is used to obtain the CFO in OFDM system. To demonstrate the efficiency of the proposed method, we compare it with other existing methods in terms of the mean square error (MSE) and estimation range. Based on the simulation analysis, the proposed method has a better accuracy result than the conventional methods at the cost of a slight decrease in estimation range when compared to the conventional methods.

The rest of this paper is organized as follows. Section 2 describes the signal model for the OFDM system. In Sect. 3, we present the concept of our method. The simulation is conducted in Sect. 4. Finally, we discuss our conclusion and future work in Sect. 5.

2 The System Model

In the OFDM transmission scheme, the data stream is split into N subcarriers and transformed to an OFDM signal by inverse fast Fourier transform (IFFT). Then, the received signal detected on the k-th subcarrier of the l-th transmitted OFDM symbol in the frequency domain, with a small CFO, is equal to

$$R_l(k) = X_l(k)H_l(k)e^{2\pi\varepsilon l N_P/N} + l_l(k) + W_l(k), \tag{1}$$

where $X_l(k)$ are the OFDM symbols during the l-th period, and N_P denotes the length of the observation training symbol. $W_l(k)$ is the additive white Gaussian noise

(AWGN) and ε is the normalize carrier frequency offset. N is the number of the subcarriers. $H_l(k)$ is the channel frequency response, and $l_l(k)$ inter-carrier interference (ICI) generated by frequency error. For the sake of simplicity, in the following derivations we neglect the ICI term since its power is very small compared with the additive noise power.

3 The Proposed Method

It is well documented that when the CFO is relatively small or the noise is very large, the difference of the rotated phases between two adjacent symbols are very small. This may result in poor estimations or in some cases give estimations of the opposite sign. If we compare the phase rotation of the current symbol with the next *S* symbol that delays *S*, the effects of noise can be reduced to some extent. It is a simple way to increase the accuracy of the *S* symbol estimation, which can be done with a differential estimator by comparing two OFDM received symbols, R_l (k) and R_{l+s-1}(k) [10].

$$\Psi(\mathrm{k}) = R_l^*(k) R_{l+S-1}(k) \tag{2}$$

In our proposed method, instead of comparing the current symbol with the next *S* symbol, we use a block of *S* observation symbols. These are grouped into two consecutive blocks, where each block has a length of *S*/2. The observation training symbols are added sequentially; the summed results are correlated as represented in the following expression:

$$\Psi(\mathrm{k}) = \sum_{\mathrm{j}=1}^{\frac{\mathrm{S}}{2}} \mathrm{R}_\mathrm{j}^*[k] \sum_{i=S/2}^{S} \mathrm{R}_\mathrm{i}[k] \tag{3}$$

In this paper, we assumed the channel is to be idle over several symbols. The Ψ(k). is then derived as in Eq. (4)

$$\Psi(\mathrm{k}) = |X_l(k)|^2 \sum_{i=\frac{\mathrm{S}}{2}}^{S} \sum_{j=1}^{\frac{\mathrm{S}}{2}} e^{j(j-i)2\pi j\varepsilon N_P/N} + \widehat{W}(k) \tag{4}$$

Therefore,

$$\Psi(\mathrm{k}) = |\mathrm{X_l(k)}|^2 \times \frac{\sin\left(\frac{2\pi \mathrm{j}\varepsilon \mathrm{N_P}\left(\mathrm{S}-\frac{\mathrm{S}}{2}\right)}{2\mathrm{N}}\right) \sin\left(-\frac{2\pi \mathrm{j}\varepsilon \mathrm{N_P S}}{2*2\mathrm{N}}\right)}{\sin\left(\frac{2\pi \mathrm{j}\varepsilon \mathrm{N_P}}{2\mathrm{N}}\right) \sin\left(-\frac{2\pi \mathrm{j}\varepsilon \mathrm{N_P}}{2\mathrm{N}}\right)} \times e^{2\pi \mathrm{S}j\varepsilon N_P/2N} + \hat{W}(\mathrm{k}) \tag{5}$$

Since ε is small, and $E\left[\widehat{W}(k)\right] = 0$, then

$$\arg\{E[\Psi(\mathrm{k})]\} = \frac{2\pi\varepsilon S N_P}{2N} \tag{6}$$

where $E[\Psi(\mathrm{k})]$ is the mean of $\Psi(\mathrm{k})$. However, in our work we proposed $N_P = N$.

$$\arg\{E[\Psi(\mathrm{k})]\} = \pi\varepsilon S \tag{7}$$

Therefore, $\hat{\varepsilon}$ which is an estimate of ε *is* derived by finding the argument of the summation of $\Psi(\mathrm{k})$ over all possible training symbols in the OFDM system,

$$\hat{\varepsilon} = \frac{1}{\mathrm{S} * \pi} \arg\left\{\sum_{k=0}^{N-1} \Psi(k)\right\} \tag{8}$$

4 Performance Analysis

The performance analysis of our method evaluated using the MATLAB simulator. We compared the simulation results of the proposed estimation method to the conventional methods. In the simulation tests, the mean squared error (MSE) is selected to be the performance metric. The AWGN is also considered when evaluating the simulation. The MSE refers to the average error within an OFDM block. The block size OFDM system parameters are chosen based on the IEEE 802.11a standard as follows: N = 64, Ng = 16, where Ng is the number of guard interval samples. The simulation tests are carried out at a different signal to noise ratio (SNR). The plots of Fig. 1 represent the first series of simulation signals with CFO that were sent over a communication channel with an AWGN.

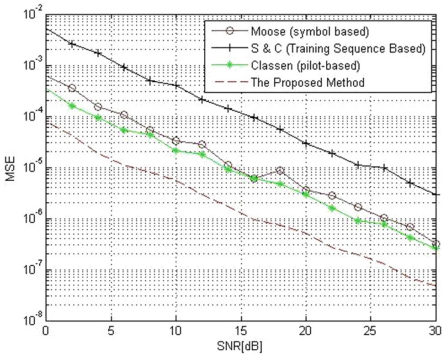


Fig. 1. MSE of CFO estimation in AWGN (S = 4, CFO = 0.02).

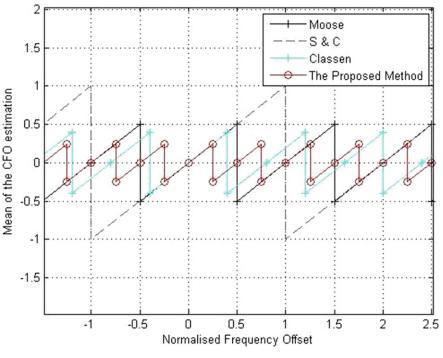


Fig. 2. Comparison for the mean of CFO offset estimation methods.

From simulation results, it could be seen that at the small CFO ($\varepsilon = 0.02$), the MSE of our proposed method estimator has superior performance compared to the conventional techniques. Between the proposed method estimator and the conventional techniques' estimators, the proposed method shows the best results; therefore, confirming the excellence of this method under $\varepsilon = 0.02$. Figure 2 shows a comparison of the estimation ranges of our proposed method to the conventional techniques of Schmidl, Moose and Classen. From the simulation results, it is found that the proposed method gives better accuracy but the estimation range is slightly decreased compare to other methods.

5 Conclusion

This paper proposed a method based on *S* symbol observation for estimating CFO using training symbols which have a good autocorrelation function during the preamble period. An analytical expression for the MSE of the frequency synchronization scheme is achieved. CFO estimation error has been considered under an AWGN channel. The proposed method has a more accurate result than the conventional techniques. The accuracy of our proposed method estimator can be improved by increasing the number of observations. However, increasing the accuracy will occur at the cost of decreasing frequency acquisition range.

The multiple input multiple output (MIMO) OFDM system is also very sensitive to CFO. Moreover, for MIMO-OFDM, there is multi-antenna interference (MAI) in the receiving antennas between the received signals. The MAI makes CFO estimation more difficult when compared to single input single output (SISO) systems. As a future work, it would be worthy to modify our method to work with MIMO-OFDM systems.

References

1. Pollet, T., Vanbladel, M., Moeneclaey, M.: BER in OFDM systems to CFO and phase noise. IEEE Trans. Commun. **43**, 191–193 (1995)
2. Peled, A., Ruiz, A.: Frequency domain data transmission using reduced computational complexity algorithms. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1980, Denver, Colorado, pp. 964–967 (1980)
3. Hirosaki, B.: An analysis of automatic equalizers for orthogonally multiplexed QAM Systems. IEEE Trans. Commun. **COM-28**, 73–83 (1980)
4. Chang, K., Han, Y., Ha, J., Kim, Y.:Cancellation of ICI by Doppler effect in OFDM systems. In: IEEE 63rd Vehicular Technology Conference, Melbourne, pp. 1411–1415 (2006)
5. Moose, P.H.: A technique for orthogonal frequency division multiplexing frequency offset correction. IEEE Trans. Commun. **42**, 2908–2914 (1994)
6. Schmidl, D.T.M.: Robust frequency and timing synchronization for OFDM. IEEE Trans. Commun. **45**, 1613–1621 (1997)
7. van de Beek, J.J., Sandell, M., Borjesson, P.O.: ML estimation of time and frequency offset in OFDM systems. IEEE Trans. Sig. Process. **45**, 1800–1805 (1997)

8. Zhou, H., Malipatil, A.V., Huang, Y.F.: Maximum-likelihood carrier frequency offset estimation for OFDM systems in fading channels. In: IEEE Wireless Communications and Networking Conference, pp. 1461–1464. IEEE Press, Las Vegas (2006)
9. Classen, F., Meyr, H.: Synchronization algorithms for an OFDM system for mobile communication. ITG-Fachtagung, pp. 105–113 (1994)
10. Liu, S.-Y., Chong, J.-W.: A study of joint tracking algorithms of carrier frequency offset and sampling clock offset for OFDM-based WLANs. In: IEEE 2002 International Conference on Communications, Circuits and Systems and West Sino Expositions, pp. 109–113. IEEE Press, Chengdu (2002)

Korean-to-Korean Translation Based Learning Contents Management System for Parents of Multi-cultural Family

YunHee Kang[1], Myung Ju Kang[2], and WooSik Kim[3(✉)]

[1] Division of Information Communication, Baekseok University, 115 Anseo-dong, Cheonan 330-704, Korea
yhkang@bu.ac.kr
[2] Triniti. Co., Ltd., Cheongwon-gun, Korea
mjkang@triniti.kr
[3] Smart Learning Korea Co., Ltd., Seoul, Korea
ceo@slkedu.com

Abstract. Language barrier, the major cause of information divide in multi-cultural families, has a close relationship with the low education level of children in multi-cultural families, which is foreseen to be an additional social problem resulted from aggravating economic inequality. Parents of multi-cultural families have restrictions in making efficient utilization of existing educational contents caused by remarkably falling behind information divide comparing with those ordinary families in terms of accessibility to digital devices and data utilization ability. In order to overcome these restrictions, it is imperative to build up a customized learning contents supporting system that provides contents appropriate to the understanding level of the learners. This paper designs the Korean-to-Korean Translation based learning contents system to dissolve the information divide in the parents of the multi-cultural families and as a result, it suggests the prototype of the Korean-to Korean Translation System to support the user customized learning contents for it.

Keywords: Multi-cultural family · Information divide · Korean-to-Korean translation · Customized learning contents

1 Introduction

In information society, the gap is getting bigger in ability between members of a society to access and utilize new knowledge and information that are important resources. This information divide is expected to lead economic inequality and social divide. It is emerging as an issue in the information society [1]. The fact finding survey in 2014 on information divide carried by the Ministry of Science, ICT and Future Planning and National Information Society Agency showed steady improvement in PC utilizing information level that evaluates information divide between ordinary people and four major alienated classes – the disabled, the low incomes, the elders, the agriculture and fishery households, the residents escaped from the North Korea, the marriage immigration women, and indicated a gradually decreasing trend of

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_141

information divide every year accordingly. The existing information divide referred to gap between those who are able to access digital information and information technology and those who are not, however, the criteria to define it is extending to the information using capability, utilizability, and efficiency.

In Korea, the number of the multi-cultural families is on the increasing trend due to increasing number of the marriage immigrations. Language barrier, the major cause of information divide in multi-cultural families, has a close relationship with the low education level of children in multi-cultural families, which is foreseen to be an additional social problem resulted from aggravating economic inequality [2, 3]. Parents of multi-cultural families have restrictions in making efficient utilization of existing educational contents caused by remarkably falling behind information divide comparing with those ordinary families in terms of accessibility to digital devices and data utilization ability. In order to overcome these restrictions, it is important to build up a customized learning contents supporting system that provides contents appropriate to the understanding level of the learners.

This paper designs a software architecture of the Korean to Korean Translation based learning contents management system to handle the information divide in the parents of the multi-cultural families and, as a result, it suggests the prototype of the Korean-to Korean Translation System to support the user customized learning contents for it.

2 Related Works

MapReduce is a program model in order to parallelize high capacity data and to do distributed processing [6, 7]. Hadoop [8] – a MapReduce middleware, being an open source project of Apache, processes data stored at HDFS (Hadoop Distributed File System), while HDFS maintains the processed result of MapReduce application. It comprises a cluster with multiple low valued servers having HDFS replication to provide higher scalability and data reliability. HDFS client requests a data block to name node. Name node maintains metadata of a data block that composes a file and provides clients with location data. Later, a client processes it by accessing data node in order to perform an actual computation on the file.

Apache Flume is an open source-based system that can deliver large amount of data securely and efficiently based on stream [9]. Flume with the distributed architecture has the benefit to make easy extension without re-setting existing server or modification even at the rapid increase in data amount. In order to process the required data for Hadoop application, an application developer stores files at HDFS or in a file unit through a program, however, it is not easy to keep the created log files at HDFS constantly as the number of servers is increasing. On account of this reason, Flume is used in the logging system and can be used to store data created under various environment in HDFS of Hadoop. Flume-based data aggregation, being in 3-tier structure, stores log data at HDFS in a single direction through a collector from agent. MapReduce program is used to analyze and process log data stored at HDFS.

3 Multi-culture Customized Learning Contents System Design

The following system guidelines are required in designing process of the multi-culture customized learning contents management system in order to suggest the customized smart learning contents appropriate to the learning surroundings and needs of individual members of the multi-cultural family.

- For the learning efficiency of the multi-cultural members, the learner's level should be evaluated first
- It is required to recommend a learning content in accordance with the level based on utilization of contents by the level of the parents of the multi-cultural family
- It is necessary to utilize a metadata standard of the learning contents for the parents of the multi-cultural family and to extend it as it required

Figure 1 shows an overall system structure of the Korean-to-Korean translation based educational contents system in order to support the Korean language learning for the parents of the multi-cultural family. The result of the Korean-to-Korean translation is maintained as a log and the log data is stored at HDFS of the big data system through the learning contents collector. The Korean to Korean translation log and the learning contents of the learning management system collected at the big data system performs recommendation through the learning contents recommendation system and then provides it to the user. This paper explains only the details limited to the sub system for configuration of the Korean-to-Korean translation system. If you have more than one surname, please make sure that the Volume Editor knows how you are to be listed in the author index.

The multi-culture customized learning contents recommendation system processes large amount of data by using HDFS – Distributed file system of Hadoop, to store the learning contents. HDFS improves performance in sequential process of large data by storing those collected data in a block unit when storing files. However, in terms of administration, it is not easy to store those created contents at HDFS steadily due to the

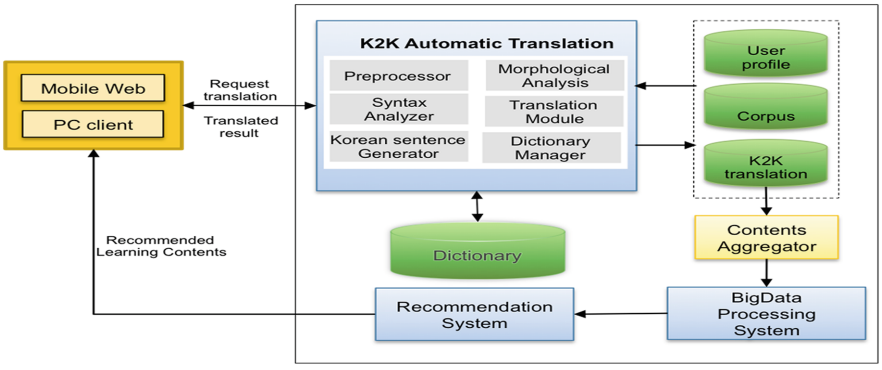


Fig. 1. System structure of Korean-Korean translation based learning contents management system

increasing number of users who create the learning contents. In order to overcome this problem, it comprises the collector using Apache Flume. Apache Flume is an open source-based system that can deliver large amount of data securely and efficiently based on stream [9]. Flume with the distributed architecture has the benefit to make easy extension without re-setting existing server or modification even at the rapid increase in data amount.

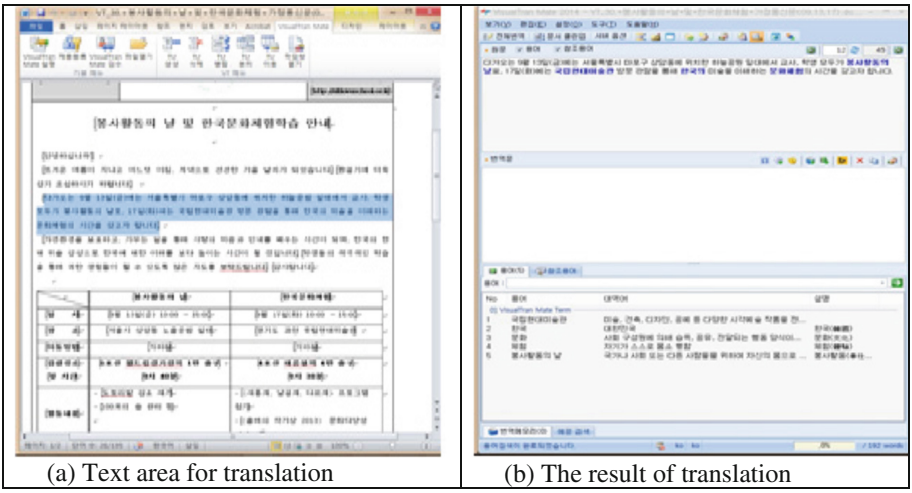
Figure 2 shows a configuration file of Apache Flume configured for operation of the collector. The relevant information has a pair of name and value. The prepared configuration file describes elements and its function. The Flume based agent stores the aggregated data at HDFS for data process like building up the index after creating a channel for the aggregated data.

The Korean to Korean translation system - the important element of the system to solve information divide in the members of the multi-cultural family – configures as on the user text input section – a field where inputting the extracted text chosen from user's input or text file – transmits the input text to the work scheduler by configuring it as a request text by a sentence unit.

```
# Name the components on this agent
a1.sources = r1
a1.sinks = k1
a1.channels = c1

# Describe/configure the source
a1.sources.r1.type = exec
a1.sources.r1.command =tail -F  K2K_Tran/log.txt
a1.sources.r1.channels = c1

# Describe the sink
a1.sinks.k1.type = hdfs
a1.sinks.k1.hdfs.path= hdfs://localhost/flume/TranData

# Use a channel which buffers events in memory
a1.channels.c1.type = memory
a1.channels.c1.capacity = 1000
```

Fig. 2. Configuration for apache flume

The Korean-to-Korean translation system selects the part of the document requiring translation at the client module as in Fig. 3(a) and transmits the selected area to the automatic translation system. The automatic translation system that received it carries out pre-processing of the sentence such as removing stop words and special characters and performs the morphological analysis and syntax analysis. A client provides the result of performing Korean language text morphological analysis on the area chosen at the editor. Figure 3(b) shows easy recommendation of Korean language for the relevant terms of Korean nouns including National Museum of Modern and Contemporary Art that is the text in the area chosen at Fig. 3(b). In this process, Korean words highlighted in blue resulted from the morphological analysis as recommendation words are obtained through researching those management/dictionary/sentence DB.



(a) Text area for translation (b) The result of translation

Fig. 3. Text area selection for Korean-to-Korean translation and the result of Korean-to-Korean translation

4 Conclusion

This paper designs the software architecture of the Korean to Korean Translation based learning contents system to dissolve the information divide in the parents of the multi-cultural families and, as a result, it suggests the prototype of the Korean-to Korean Translation System to support the user customized learning contents for it. The designed Korean-to-Korean Translation based learning contents system is expected to solve the information divide issue in parents of the multi-cultural families. Later, it is scheduled to build up a user-customized learning contents supporting system that provides an appropriate contents to the understanding level of a learner.

Acknowledgements. This material is based upon work supported by the Ministry of Trade, Industry & Energy(MOTIE, Korea) under Industrial Technology Innovation Program. No.10059094, 'Technology development of smart learning and cultural exchange for multi-cultural members'.

References

1. van Deursen, A.J.A.M., van Dijk, J.A.G.M.: The digital divide shifts to differences in usage. New Media Soc. **16**(3), 507–526 (2013)
2. Albulut, Y., Cardak, C.S.: Adaptive educational hypermedia accommodating learning styles: a content analysis of publications from 2000 to 2011. Comput. Educ. **58**, 835–842 (2012)
3. Verbert, K., et al.: Context-aware recommender systems for learning: a survey and future challenges. IEEE Trans. Learn. Technol. **5**(4), 318–335 (2012)

4. Adamopoulos, P.: What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. In: Proceedings of the 34th International Conference on Information Systems, ICIS (2013)
5. Jang, S.H.: A comparative study on the development of K-MOOC platform. J. Platform Technol. **4**(1), 33–28 (2016)
6. Dean, J., Ghemawat, S.: MapReduce: a flexible data processing tool. Commun. ACM **53**, 72–77 (2010)
7. Kang, Y.-H.: Construction of a MapReduce application running on twister in cloud computing environments, futuregrid. J. KIIT **9**(4), 147–154 (2011)
8. http://hadoop.apache.org/
9. http://flume.apache.org/

A Distributed Survey Automation Based on a Customizable Form Template

Jaekwon Lee[1], Kisub Kim[1], Jang-Eui Hong[2], and Woosung Jung[3(✉)]

[1] Department of Computer Engineering, Chungbuk National University, Bd. E8-1, Cheongju, South Korea
{exatoa, falcon}@cbnu.ac.kr

[2] Department of Computer Science, Chungbuk National University, Bd. E8-10, Cheongju, South Korea
jehong@cbnu.ac.kr

[3] Seoul National University of Education, Graduate School of Education, R&C Block, Seoul, Republic of Korea
wsjung@snue.ac.kr

Abstract. The survey is a common approach to checking the validity and reliability of specific opinions, where a large enough number of samples is required to provide its statistical significance. However, if the survey data is too big for one person to handle and takes too much time, the questionnaire set should be divided into small pieces to reduce the respondent's burden. In this paper, we propose a novel and practical approach to distribute survey data and generate questionnaires automatically based on customizable form templates. The separately collected results are merged into predefined repositories for further research. We apply the proposed method to evaluate the quality of bug reports. The experimental results show that it is relatively more effective than the manual process.

Keywords: Survey · Questionnaire · Data mining · Automation · Genetic algorithm

1 Introduction

Data mining is used to discover large amounts of data that are difficult to obtain superficially from raw data. In the software engineering field, it also helps developers to build software systems effectively and reduce the overall cost of development with useful information extracted from project artifacts.

However, there is some information that cannot be extracted from existing artifacts, for example, the opinions of experts. To gain those kinds of information, researchers have used the survey. The questionnaire is a common survey instrument used for checking the validity and reliability of specific opinions where a large enough number of samples are required to provide its statistical significance. Recently, Zimmermann [1] used a questionnaire-based method to investigate the features of good bug reports, and Pham [2] also used the approach to investigate how clarify effects of the GitHub affects to developer's testing behavior in the project. The questionnaires were also used

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_142

to evaluate the accuracy of the proposed approaches [3, 4]. Though those researchers could easily obtain feedback from one questionnaire, it is not suitable for the cases in which different questionnaires need to be deployed to each respondent. Furthermore, distributing data and generating, deploying, and merging questionnaires are a time consuming task. The more survey data we have, the more time and cost will be consumed.

In this paper, we propose an automated approach for distributing data, generating questionnaires, and merging the results following specified constraints. The system based on the approach receives data, conditions for distribution, and a template for the questionnaire as the inputs. Then it generates questionnaires the using Genetic Algorithm (GA) and Google Apps API, and it finally merges the results from the targets. We performed experiments to show the effectiveness of our approach by applying it to an evaluation of quality for duplicate bug reports. The results confirm that the system can generate questionnaires suited for 10 bug report set and three user-defined constraints and merge the deployed responses appropriately. We also verified that the reports that were evaluated as high quality have code examples.

2 The Automation Process of Survey Using Multiple Forms

In this paper, we propose a process that consists of three steps to create forms, deploy them to the respondents, and merge the results as shown in Fig. 1.

As the first step, using the GA, we divided the survey data by the number of respondents and then generate questionnaires about the each survey data. We used Google Form, which is useful for deploying questionnaires and arranging the data of

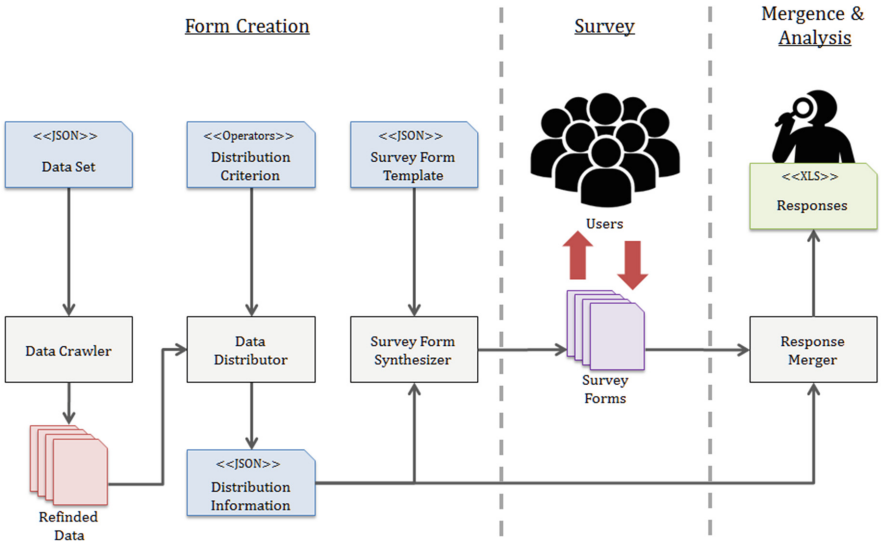


Fig. 1. The automation process of creating forms and merging data

responses, to create the questionnaires. We needed K questionnaires since each respondent should get different questionnaires. In this step, we used Google Apps Script for automation. However, Apps Script tends to be slow (it takes approximately one minute to create a form) and the processing time is five minutes, so we used Apps Script at minimum works and the rest of the system was developed in python.

Delivering questionnaires that were created in the previous step to the target respondents and collecting results from them was the second step. This can be accomplished not only by opening the links according to the characteristics of the questionnaires but also by deploying those to specific groups of people. The key is choosing the appropriate group in order to control the constraints needed for the research.

The last step was merging and analyzing the performed questionnaires. The responses of Google Form from respondents were stored in each survey file. To reduce the cost for data integration, all of the survey files were also integrated into one final file using Google Apps Script. Because the each survey file had no information about the respondent, Distribution Information was used to map the distribution information and responses. Researchers finally analyzed the integrated data to extract the information they needed.

The following sections explain the modules of the generation process.

2.1 Crawling Data

In the first step, the *Data Crawler* collected and refined the original data based on the entered *Data Set*. The *Data Set* contained the list from the selecting process for the needed data. Because it had to be constructed separately according to the research area, we assumed that we should collect duplicate bug reports from various domains. In this case, the input data was the list of bug reports for each domain *MATH* and *LANG* as JSON formats and is shown in Fig. 2.

The *Data Crawler* downloaded and stored the bug reports from the online repositories based on the input data, and we eliminated the extra information, such as comment, carbon, and copy (cc), which was not included in the initial state of the

```
{
    "MATH":{
        list:[[295, 294], [1280, 1252], [846, 801], [886, 611], ...],
        url : "https://issues.apache.org/jira/si/..."
        type: "JIRA"
    },

    "LANG":{
        list:[[414, 346], [888, 889], ...],
        url : "https://issues.apache.org/jira/si/..."
        type: "JIRA"
    },
    ...
}
```

Fig. 2. The example of the bug report set

report. It is a control condition and makes the same circumstance when a developer faces the bug report for the first time. After collecting and refining was done, the system transferred data containing additional data like the url of the bug report needed to create the questionnaires for the *Data Distributor*.

2.2 The Target Data Distribution with the Genetic Algorithm

We wanted to make a distribution of N target data items to K respondents. The distribution problem is an NP problem because it is hard to design the algorithm in polynomial time, so we should search all possible solutions. The GA is one of the meta-heuristic search techniques that mimics the principle of natural selection. The algorithm must define the chromosomes and the fitness function and works with many operations such as crossover, mutation, selection, and replacement. The following paragraphs describe the structure of chromosomes, operations, and the parameters.

Chromosome Representation. A chromosome is an integer string and each one of the element in the string is mapped to each item of data. The value of an element in a chromosome is the number of the group that will be assigned data, i.e., If N data were distributed to K respondents, the chromosome size would be N and each element would have a value between 1 to K. For example, if the first value of a chromosome is 3, that means the first data item is assigned to respondent number 3.

Fitness Function. This function derives an objective chromosome controlling direction of evolution of the chromosome. In this study, we use the average of errors from one chromosome as the fitness measure and define number 0 as the best solution. Figure 3 shows the process of calculating the fitness score in the case of distributing six items of data to three questionnaires. When we assume a chromosome has values like (a), the solution can be (b). There are two errors in each group, 2 and 3, as shown in (c). Finally, we obtained 0.6667 scores that calculated the average of error counts.

Evolution Operations. We applied a steady-state GA considering elitism to our system. The algorithm started with a 1,000 population and each chromosome was filled with random integers. For the next generation, we chose two parents using roulette selection and created offspring with uniform crossover. Then the offspring was mutated in specific probability. The probability decreased steadily in every generation. Finally, the parent with a low fitness was replaced immediately with the offspring. Those operations maintained the diversity of the population and prevented rapid convergence. The population evolved until the system produced the best solution.

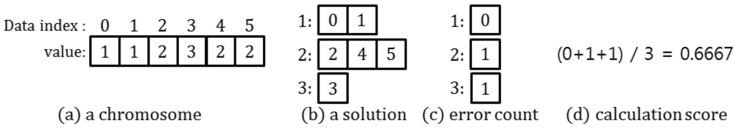


Fig. 3. The calculation process of the score in fitness function

```
{
    "subjects": [
        [[0, 1], [2, 10], [4, 15], [9, 3], [11, 7]],
        [[10, 2], [5, 8], [6, 12], [14, 20], [21, 30]],
                        ...
        [[44, 21], [2, 10], [6, 12], [18, 25], [29, 27]]
    ],
    "bugs":[
        {"id":195, "domain":"MATH", "url":"http://.../195.html"},
        {"id":443, "domain":"MATH", "url":"http://.../443.html"},
                        ...
        {"id":392, "domain":"COLLECTIONS", "url":"http://.../392.html"}
    ]
}
```

Fig. 4. An example of the distribution information for evaluating duplicate bug reports

The Result. The *Distribution Information*, the result of GA, was represented by the JSON format. Figure 4 is an example of the distribution information for evaluating duplicate bug reports. Each line of subjects, having a list of five pairs, was assigned to

```
{
"form":{
    "output":"0B7qm6CAi3HmhNkh2WFR1ZHNkNGM",
    "filename":"duplication_%02d",
    "title":"The Quality of Duplicate Bug Report",
    "desc":"",
    "progress":true
},
"pages":[
    ...
    {
        "title":"%s : [%d -- %d]",
        "desc":"The following bug reports are duplicate report.
                Read carefully and answer the questions",
        "items":[
            {"type":"Section", "title":"Bug Report", "desc":""},
            {"type":"Section", "title":"",
             "desc":"[%s-%d] : %s \n\n[%s-%d] : %s \n\n"},
            {"type":"Section", Questions", "desc":""},
            {"type":"Scale",
             "title":"1. How quality of bug report [%s-%d]?",
             "bounds":[1,5], "labels":["low","high"],
             "required":true},
            {"type":"Scale",
             "title":"1. How quality of bug report [%s-%d]?",
             "bounds":[1,5], "labels":["low","high"],
             "required":true}
        ]
    },
    ...
]
}
```

Fig. 5. An example of a form template

each respondent and bug containing all the information of the bug reports. The numbers used in subjects were the indices of bugs and the information in bugs was the target data that was used to generate questionnaires.

2.3 Form Creation

We used Google Form to create the questionnaires. The *Survey Form Synthesizer* generated the final form combining *Distribution Information* and the *Form Template*. The *Synthesizer* utilized Google Apps Script to make a form but as we already mentioned, Apps Script has limitations on performance and execution time. Therefore, we defined functions to make one form instance in Apps Script and make the *Synthesizer* call it.

Figure 5 is an example of the *Form Template* written in JSON format. The template was composed of two factors, *form* and *pages*. The *form* included the information, the output folder, file name, form title, form description, and other options. The *pages* consisted of the page list, which contained page title, description, and items. The item had various types of components like Section, TextBox, Scale, CheckBox, etc. They had different properties depending on the type except for the title and description. The detailed properties are described in a Google Form API document [6].

3 Applying the Approach for Bug Report Quality Evaluation

In order to verify the proposed method, we produced questionnaires using our system. Then we deployed them and merged the responses from respondents. For this study, we used duplicate bug reports to determine which features were useful for understanding the bug. Because duplicate bug reports have two different descriptions for the same bug, it is suitable to evaluate the usefulness of features or explanation.

3.1 Input Data

To create the questionnaires, we first selected four open source projects, Apache commons-collections, commons-io, commons-lang, and commons-math. These projects prevent getting a result that is subordinated to a specific project characteristic. There were 127 pairs of bug reports, and we filtered duplicate bug reports that had patches in the version control system while those are not complementing each other, and then we chose 10 pairs randomly. Second, we defined the following list as the distribution conditions that would reduce the respondents' load, make them not to be familiar with the project, and increase the reliability of the responses.

- Only 5 pairs of bug reports were assigned to one questionnaire.
- One questionnaire should be assigned at most 3 pairs of bug reports.
- Each bug report should be evaluated at least 3 times.

We defined the questions and form template as in Fig. 5 (see Sect. 2.3).

3.2 The Results of the Study

As we needed to make three evaluations for 10 pairs of duplicate bug reports respectively, the data for the survey was 30 pairs including duplicate pairs. For this reason, we created six types of questionnaires that contained five pairs of reports each. Table 1 shows a summary of responses from six respondents. The columns for scores represent the results of the evaluation using the five-level Likert scale and the average of three scores are positioned in Avg. column. We inferred what properties affect to the quality of a bug report and analyzed the correlation between the scores and properties of the bug reports. For example, there were two points difference on average scores for duplicate bug reports, IO:179–339 and Lang:464–468. One had a *short description* and *source code example* and the other had a *short description* and *link*. Since the first one was much higher on average than the other, it finally revealed that the *code example* improved the quality of the bug report. Math:418–1112 scored three points as an average score with a short description and link property. This pair helped us to evaluate the fitness of the score compared with another pair that had similar properties. However, since the summary was the extracted result from the minimum number of responses, the reliability of the survey was low. To improve this, we should get more responses to each questionnaire.

Table 1. The summary of response data from respondents

Domain	Report		Score 1		Score 2		Score 3		Avg.	
	A	B	A	B	A	B	A	B	A	B
Collections	525	577	3	4	4	2	4	2	3.7	2.7
Collections	583	580	2	4	3	3	2	3	2.3	3.3
IO	79	132	1	4	5	3	2	4	2.7	3.7
IO	197	339	3	5	1	3	2	5	2.0	4.3
Lang	464	468	5	3	4	1	4	3	4.3	2.3
Lang	477	509	4	1	5	4	5	4	4.7	3.0
Math	294	295	2	3	5	5	5	5	4.0	4.3
Math	418	1112	2	3	3	3	4	3	3.0	3.0
Math	781	813	3	4	2	4	2	4	2.3	4.0
Math	1203	1234	3	2	4	4	3	3	3.3	3.0

4 Conclusion

In this paper, we proposed an automated approach that deploys a large amount of survey data to target respondents and provides the integrated results of all questionnaires. The system operates with given input including information on the survey data, distribution constraints, and template for the questionnaires. It crawls and manufactures data and then divides data among questionnaires using the GA. We used Google Forms for generating questionnaires. Finally, the responses from the respondents were merged and analyzed. This system will be useful to the public and researchers who want to

collect needed data, especially where there is a large amount of survey data or it takes too much time to merge the results.

We recognize that there are some weaknesses involved in the model for the system. Because the GA is suitable for finding an approximate solution, if the conditions become more complex or the amount of survey data increases, it is more difficult to obtain accurate results. For future work, we will use a large amount of survey data with complex conditions and overcome the limitations. Furthermore, we will use this system to judge the features that we should treat as a valuable when we try to find the location of fault from bug reports.

Acknowledgments. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIP) (No. 2015R1C1A1A01 054994, No. 2014M3C4A7030505).

References

1. Zimmermann, T., Premraj, R., Bettenburg, N., Just, S., Schroter, A., Weiss, C.: What makes a good bug report? IEEE Trans. Softw. Eng. **36**, 618–643 (2010)
2. Pham, R., Singer, L., Liskin, O., Filho, F.F., Schneider, K.: Creating a shared understanding of testing culture on a social coding site. In: 35th International Conference on Software Engineering, pp. 112–121 (2013)
3. Bruch, M., Monperrus, M., Mezini, M., Briand, L.: Learning from examples to improve code completion systems. In: 7th joint meeting of the European Software Engineering Conference and the ACM SIGSOFT Symposium on the Foundations of Software Engineering, pp. 213–222 (2009)
4. Carreno, L.V.G., Winbladh, K.: Analysis of user comments: an approach for software requirements evolution. In: 35th International Conference on Software Engineering (ICSE), pp. 582–591 (2013)
5. Cavicchio, D.J.: Adaptive search using simulated evolution. Ph.D dissertation, University of Michigan (1970)
6. Google Apps Script Forms API. https://developers.google.com/apps-script/reference/forms/

Mobile Agent Oriented Service for Offloading on Mobile Cloud Computing

HwiRim Byun, Boo-Kwang Park, and Young-Sik Jeong(✉)

Department of Multimedia Engineering,
Dongguk University, Seoul, Korea
{hazzzly, pbg0517, ysjeong}@dongguk.edu

Abstract. With the performance development of mobile devices, applications requiring high computing power are increasing. Therefore, the need for mobile cloud computing (MCC) is emerging to improve the computing. MCC can enhance performance for the required resources by integrating, managing and using the resources of many mobile devices. MCC models that are being serviced now use various offload processing methods, but no offload method considering the characteristics of mobile device has been defined yet. In this paper, offload methods that transfer the tasks of mobile devices from the MCC environment to the cloud are classified into Client-Server Communication, Virtual Machine Migration, and Mobile Agent, and Mobile Agent is subdivided into Client Mobile Agent and Server Mobile Agent. Thus, a total of five offload methods are defined. Instead of the offload methods that do not consider the mobile environment, the Mobile Agent Oriented Service (MAOS) is proposed as a new offload method.

Keywords: Mobile device · Cloud computing · Offload · Computing resource · Mobile cloud computing · Distributed computing

1 Introduction

Various mobile devices such as smartphone, tablet PC, and smart watch are being established as the fastest convenient communication tools in the present era. However, their limited performance due to the small size for mobility interferes with the enhancement of service quality. Therefore, mobile cloud computing (MCC) is being actively researched to improve performance by sharing resources such as storages and processors in one network of many mobile devices as shown in Fig. 1. MCC has been in the spotlight because it allows mobile devices to have a higher performance than the existing devices at low cost. Therefore, it has emerged as the next-generation computing infrastructure. Even though various services are being provided for the management and resource efficiency aspects of MCC, classifications of offload methods have not been clearly defined yet.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_143

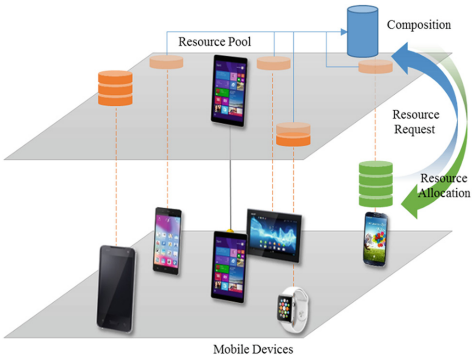


Fig. 1. Workflow of MCC

2 Classification of Offload Methods

Offload refers to the transmission of specific computing jobs from a mobile device to an external platform. The limitation of existing resources is overcome by performing jobs using the resources of a cloud rather than the resources of the mobile device itself. As a result, it can provide virtual mobile performance with a faster processing speed than the existing method. Figure 2 shows the offload classification proposed in this paper Fig. 1.

2.1 Client-Server Communication

In this method, mobile devices directly access surrogate servers comprising a cloud through the network. This means the direct connection relationship of client and server. This method offers stable, excellent processing capacity for the common services of client and server. In particular, the more the number of clients simultaneously connected to the server is, the greater the expected value of the available resources becomes. However, this method has problems in terms of cost and management because a single server must provide resources to all the connected mobile devices. Furthermore, the services must be installed in the participating devices in order for them to call other services.

2.2 Virtual Machine Migration

In this method, a user environment of mobile devices is constructed through OS virtualization above the physical hardware layer of the server. Clients can directly install and use required services on the OS and, as a result, the server can respond to clients' calls for various services. When data migration between servers is required, the memory image of the virtual machine is sent from the origin server to the target server to implement uninterrupted services of OS and applications. This method has an advantage in that no additional source code is necessary to provide new services because it uses the OS. Furthermore, the boundary of the virtual machine provides

excellent security. However, the integration time of the VM can cause a delay. Virtual Machine Migration can be divided into two methods depending on the matching method of the client and the virtualization OS: single virtual machine server if the virtualization OS and client are matched as one to many, and multi virtual machine server if they are matched as one to one.

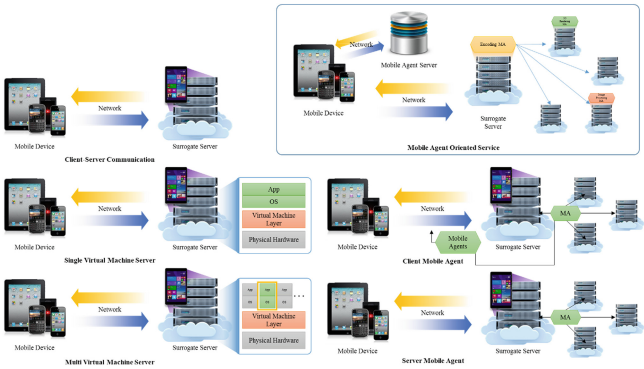


Fig. 2. Classification of offload methods

2.2.1 Single Virtual Machine Server

In this case, the server has one virtual OS and all clients access the same virtual OS. As the clients can get involved in the jobs of other clients, security reinforcement effect by the virtual machine boundary cannot be expected. However, clients can be flexibly added and removed when multiple clients process the same job in a distributed manner.

2.2.2 Multi Virtual Machine Server

The server has multiple virtual OS's and every client is matched one-to-one with an individual virtual OS. This method has excellent security because each virtual machine is independent and clients cannot access the jobs of other clients due to the virtual machine boundary. However, Multi Virtual Machine Server has virtual machine overlay compatibility problem and one server cannot respond to many clients.

2.3 Mobile Agent

The client collects required resources through a mobile agent which is an independent program running on the server. Mobile agents are sent from a server to another connected server. The server recognizes mobile agents that need resources and provides them. This method is excellent for real-time dynamic deployment. As mobile agents run independently from the server, they are ideal for an environment that has a high possibility of network disconnection such as mobile devices. However, this method has security vulnerability because agents that have information about services access many servers.

Mobile agents can be divided into client mobile agent, which is sent from the client to the server and performs processes, and server mobile agent, which runs on the server and immediately responds to clients that require resources.

2.3.1 Client Mobile Agent

Programs written in script language are offloaded from the mobile device to the server and runs on the server. The program itself is sent unlike the conventional communication in which data is sent. The server collects and provides resources that clients need. This method can flexibly respond to the requests of clients because the server and client do not need to have applications that perform the same services.

2.3.2 Server Mobile Agent

The mobile agent collects resources on the server before clients request them and can immediately provide the collected resources when the clients request them. Although dynamic deployment is possible because the agent is always running, resources can be wasted if the resources requested by client are different from those collected by the existing mobile agent running on the server (Table 1).

Table 1. Comparison of offload methods according to the MCC environment

Offload type	Characteristics	Comparison with MAOS
Client-Server communication	-Most effective for single service processing. -Flexible response to various mobile application environments is difficult.	-MAOS can respond to various applications and environments. It depends on the type of mobile agent.
Single virtual machine server	-Advantageous for single process collaboration. -Data processing issues such as personal information	-MAOS does not require initial settings for cloud infrastructure unlike the Virtual Machine Server.
Multi virtual machine server	-Advantageous for data processing such as personal information -Inefficient for mobile cloud that has an intermittent access pattern.	-MAOS has no issue of the virtual machine overlay compatibility.
Client mobile agent	-Can be used even in unstable network environment. -Network data needs to be used for agent transmission.	-MAOS prevents the unnecessary data transfer of clients because the server has mobile agents in them.
Server mobile agent	-High processing speed can be attained. -Limited to the services of mobile agents that are provided by the server.	-MAOS has a wide range of services because many servers can have specific mobile agents.

The MCC environment must choose an offload method after sufficiently considering network instability. However, this issue is not considered except for mobile agents, and mobile agents also have an issue in that they must use a large network bandwidth. The Mobile Agent Oriented Service (MAOS) can be used flexibly on mobile devices considering network instability.

3 Mobile Agent Oriented Service (MAOS)

In this paper, the new offload model Mobile Agent Oriented Service (MAOS) is proposed for the network environment that is interrupted intermittently with the exclusion of large volume transmission.

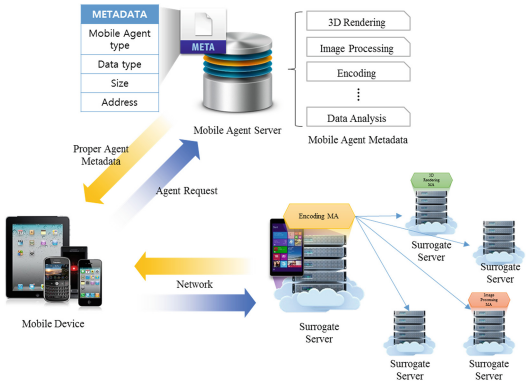


Fig. 3. Proposed MAOS model

The MAOS model is shown in Fig. 3 Some of the servers comprising the cloud have mobile agents installed on them to process data. These mobile agents can be of different types and have different data to process. The mobile agent server collects information about mobile agents that each server has. Therefore, it can indicate which server has a specific mobile agent that is needed by mobile devices. The server sends mobile agents to other servers for distributed processing.

(1) The mobile device communicates with the mobile agent server to obtain information about the mobile agent that it wants to use. The required type of mobile agent, the data format to be processed, data size, and the network address of the mobile device are transmitted using metadata.
(2) The mobile agent server organizes information about the most appropriate agent among the mobile agents that distributed servers have based on the metadata and returns the information to the mobile device.
(3) The mobile device connects with the server that has the corresponding mobile agent based on the information received from the mobile agent server and sends the data to the server.

(4) The server having the mobile agent organizes a resource pool to process the received data with, divides the data and sends it to the available server together with the mobile agent.
(5) The processing result of each server is sent to the mobile device by the mobile agent and a merging job is performed.

The data that can be processed varies by the type of mobile agent that each server has. The servers comprising a cloud do not have the same mobile agents, but install them as needed to perform jobs. The cloud server is constructed in the form of Infrastructure as a Service which provides data processing resources to mobile devices. Furthermore, it also has the form of Software as a Service because the servers comprising the cloud share mobile agents as well as resources through the Infrastructure as a Service.

4 Conclusions

This paper proposed the classification of five offload models based on the existing studies and analyzed the characteristics of each model when applied to the MCC environment. The MAOS model which is ideal for the MCC environment was presented based on this research. The MAOS processes all jobs regarding distributed processing on the cloud after one data transfer. Therefore, it can cope with irregular network interruptions, which is a characteristic of mobile devices. Based on the results of this study, more studies on the definition and design of detailed functions of the MAOS will be conducted in the future.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2014R1A1A2053564). And also This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) supervised by the IITP (Institute for Information & communications Technology Promotion), And also This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government(MSIP) (No. R7120-16-1006, Resource Integration Management Solution with On-premises Legacy for Intra-Cloud).

References

1. Dinh, H.T., Lee, C., Niyato, D., Wang, P.: A survey of mobile cloud computing: architecture, applications, and approaches. Wirel. Commun. Mob. Comput. **13**(18), 1587–1611 (2013)
2. Guan, L., Ke, X., Song, M., Song, J.: A survey of research on mobile cloud computing. In: Proceedings of the 10th IEEE/ACIS International Conference on Computer and Information Science, pp. 387–392. IEEE Computer Society (2011)

Unstructured Data Service Model Utilizing Context-Aware Big Data Analysis

Yonghoon Kim[1] and Mokdong Chung[2(✉)]

[1] Pukyong National University, 45, Yongso-Ro, Nam-Gu, Busan 48513, Korea
kimyhjava@pukyong.ac.kr
[2] Department of Computer Engineering, Pukyong National University, Busan, Korea
mdchung@pknu.ac.kr

Abstract. Recently, the analysis of the structured data and the unstructured data in the various domains such as transportation and health care is essential. Especially, it requires intelligence and/or needs analyzing unstructured data, which accounts for a large proportion from big amount of information for the purpose of a more accurate information. Therefore, in this paper, we propose a unstructured data service model which provides a reliable and more precise information by using unstructured data analysis, and we try to show an example of application of the transportation system. We use KO-NLP, LSA, FCM Clustering, and machine learning techniques. We expect the proposed model might be used in the complex domains including too much unstructured data.

Keywords: Latent semantic analysis · Fuzzy C-Means Machine learning

1 Introduction

Development of information technology has made possible the accumulation of large amount of data in the internet, and increases social and political interest in the development of Big Data analysis. However, the domestic Big Data analysis relied on the foreign companies may be a serious issue, and there may be data exclusivity in the real world.

Big Data analysis including structured and unstructured data is essential element. Up to date, however, reliable and meaningful research on Big data analysis is done for the structured data only, and this limitation reduces reliability of the analysis in a certain domain whose unstructured data portion is over more than 80 % of corporate data [1]. This means a significant number of important information cannot be utilized in the real world in terms of Big Data Analysis [2].

In this paper, with respect to the problem with unstructured data mentioned above, we use a statistical method for unstructured data analysis, Latent Semantic Analysis (LSA) to solve the problem and formalize unstructured data, and also utilize machine learning technique in terms of Fuzzy C-Means (FCM) algorithm.

The remainder of this paper is organized as follows. Section 2 surveys the related work and the theoretical backgrounds. Section 3 introduces the proposed Context-Aware

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_144

Analyzer. In Sect. 4, we apply the proposed analyzer to a comparative analysis of the experimental result of the amount of rainfall. Finally, Sect. 5 presents our conclusion and the direction of our future research.

2 Related Work

2.1 Korean Natural Language Processing

Tagging is a part of speech given to the corpus or the one thing selected parts of speech from the ambiguous results of morphological analysis. This research presents the morphological analysis utilizing HanNanum Ko-NLP Java version which was Semantic Web Research Center (SWRC) at Korea Advanced Institute of Science and Technology.

2.2 Latent Semantic Analysis

The mathematical foundation of LSA is the Vector Space Model (VSM), an algebraic model for representing documents as vectors in a space where dictionary terms are used as dimensions [3].

LSA is related to the analysis of the hidden meaning in a sentence. Deerwester et al. (1990) proposed an elegant way to improve similarity measurements with a mathematical operation on the term-document matrix, X based on Linear algebra. Deerwester, Landauer and Dumais (1997) focused on the document similarity rather than the word similarity. Document similarity is treated a main issue in Latent Semantic Indexing (LSI) while the words of similarity a main issue of LSA. They use a statistical method for the Vector Space Models (VSMs) based on Singular Value Decomposition (SVD) [4].

2.3 Fuzzy C-Means Clustering Algorithm

Clustering is a method of classifying given data by comparing them to the pre-fixed class until a class closest to the fixed class is found. The Fuzzy C-Means (FCM) clustering algorithm is a data-classifying algorithm that uses the Fuzzy division technique which classifies data points according to the membership degrees. The membership function U of the FCM clustering algorithm has elements that have values ranging between 0 and 1, and the sum of membership values for the data set is always 1 [5, 6].

3 Design of Context Aware Analyzer

Figure 1 is a Context-Aware Analyzer (CAA) for processing structured and unstructured data.

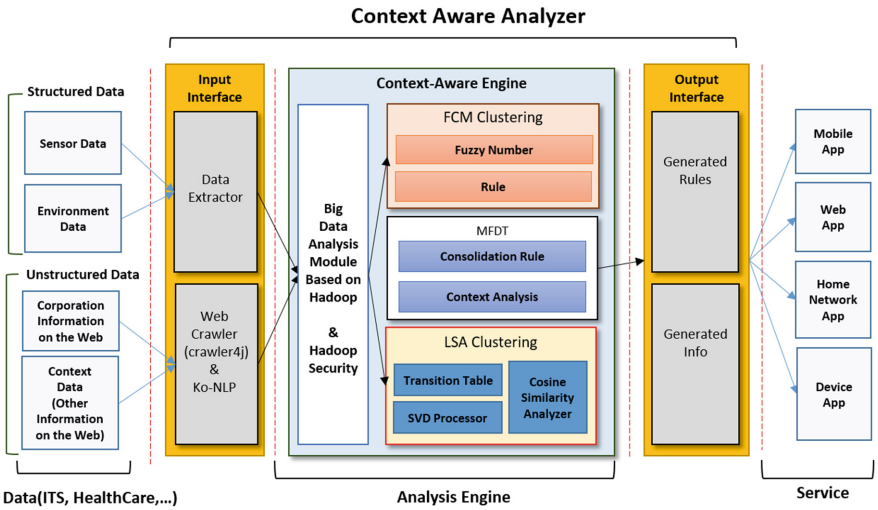


Fig. 1. Context Aware Analyzer

3.1 Input and Output Interfaces

Input Interface consists of structured and unstructured module, and structured data is stored in the Big Data store, unstructured data is stored in the forms of morpheme unit after the data collection by the Web Crawler using the Korean Natural Language Processing (KO-NLP). Output Interface contains the generated rules and information owing to the processing of Context-Aware Engine.

3.2 Context-Aware Engine

3.2.1 Context Analysis

The context analysis is speculation about the circumstances consisting not limited to word and phrases, sentences, and applies additional comparative review and test for correct conclusions by introducing Multivariate Fuzzy Decision Tree (MFDT) in the proposed model [7].

The context analysis calculated by applying Bayesian networks and Hidden Markov Model is useful since it may have learning observed evidences shown in the overall algorithm in Fig. 2.

3.2.2 Analysis Engine

Analysis engine consists of Fuzzy C-Means (FCM) clustering, Multivariate Fuzzy Decision Tree (MFDT) and Latent Semantic Analysis (LSA) clustering modules. FCM computes maximum, minimum, and center values using the FCM algorithm in the stored data, and it calculates the membership function according to the value. And the unstructured data is constructed using the transition table of LSA clustering, is decomposed into Singular Value Decomposition (SVD) Processor, is compared using a cosine similarity analyzer, and finally is transferred to FCM clustering module.

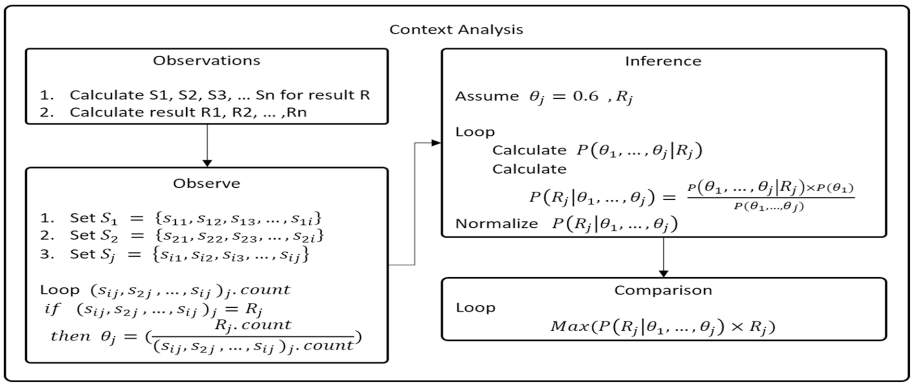


Fig. 2. Context Analysis Algorithm

The clustered data from FCM clustering combines multiple Fuzzy rules into singular rule using MFDT, reviews all environmental factors utilizing control manager, is stored in the forms of generated rules, and is finally serviced as estimated information to the input value of the user.

4 Implementation and Evaluation

4.1 Implementation of Structured and Unstructured Data Clustering

In this paper, we applied the proposed context-aware analyzer for intelligent transport systems, where the traffic data were collected on travel time, traffic volume, and the amount of rainfall in Busan and Gimhae City, Korea.

We used the public data provided by the Korea Highway Corporation (travel times 5,543, traffic volumes 8,606, rainfalls 720), and used the event information data provided by the Korea Tourism Organization (consists of documents 17 and words 2,972), and we performed experiments to predict the traffic volumes and travel times by using the FCM clustering and LSA clustering, used a MatLab and Java to this analyzer. Events information for the construction of unstructured data was extracted from the site, http://korean.visitkorea.or.kr, and we have inserted two-unrelated data from the daily newspaper [8] to distinguish between construction and event information when inserting large quantities of documents.

4.2 Evaluation of Structured and Unstructured Data Clustering

We predict the traffic volume of one week by collecting real-time traffic information. Our goal is to maximize the accuracy of the traffic and the experimental results are shown in Table 1, where the accuracy is 101.45 %, and 1.45 % have been exceeded.

Also we have compared our proposed system with the typical fuzzy system, where overall average is similar. But there is significant difference in the standard deviation. This result shows that our proposed system is more reliable and accurate.

Table 1. Proposed System and one by one week means of Actual Data Comparison

Weekday	Proposed fuzzy system	Typical fuzzy system	Actual data (Number of vehicles)		Comparison	
			1w	2w	Proposed. F	Typical.F
SUN	34,319	37,416	28,821	34,382	96.86 %	88.85 %
MON	34,348	35,350	30,251	33,125	96.42 %	93.69 %
TUE	33,991	34,865	32,302	31,990	100.00 %	97.49 %
WED	34,749	33,978	33,360	35,154	102.54 %	104.87 %
THU	34,537	35,156	34,805	35,681	104.90 %	103.05 %
FRI	36,581	37,835	38,470	39,237	107.92 %	104.34 %
SAT	38,913	41,510	40,742	37,180	101.48 %	95.13 %
Mean	**35,348**	**36,587**	**34,107**	**35,250**	**101.45 %**	**98.20 %**
Variance					0.15 %	0.32 %
Standard deviation					**3.84 %**	**5.65 %**

We could infer that the impact of external vehicle flows increased by Lotte Gimhae Water Park sale event (07/07/2016-7.9) and open ceremony (15/07/2016). To increase and decrease the problematic data, we need to use unstructured data as well as structured data. If we apply the proposed model to the unstructured data analysis, it is expected to predict the results in more accuracy and reliability. The implementation is going on for the unstructured data analysis.

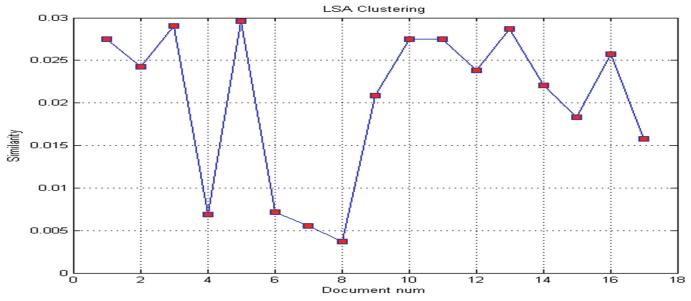


Fig. 3. Example of 17 cases of classified documents

Figure 3 is showing an affinity for the article, and the documents 1–3, 5, 9–17. The documents 9–17 is related the events information, and the documents 1–3 and 5 is related the construction information, and the documents 4 and 6–8 is related the daily

paper. The documents 4 and 6–8 was able to confirm that the low degree of similarity as a document that is not to do with the event information and construction information in a daily newspaper [8] article.

We tried to test by inserting an additional document event as shown in Fig. 3. We confirmed that it was possible to accurately classify a large number of additional documents. The document classification was done for applying the correlation and association analyses configuring the three-dimensional matrix based on the mutual link between words, and later it could be applied to machine learning by the three-dimensional coordinates.

5 Conclusion and Future Research

In this paper, we proposed a Context-Aware Analyzer (CAA) which could handle unstructured data as well as structured data in the existing structured data dominated system. The CAA used FCM clustering for structured data analysis and used LSA for unstructured data analysis which could guarantee more reliable service model in Big Data analysis.

Unstructured data analysis is in progress and furthermore we try to extend the scope of research to the diverse unstructured data analysis.

Acknowledgement. This work was supported by a 2016 Research Grant from Pukyong National University.

References

1. Unstructured Data and the 80 Percent Rule, Seth Grimes. http://breakthroughanalysis.com/2008/08/01/unstructured-data-and-the-80-percent-rule/
2. Park, J., Park, K., Lee, Y.: Big data as a new technology paradigm. Sci. Technol. Policy, 17–30 (2013)
3. Evangelopoulos, N.: Latent semantic analysis. Wiley Interdisc. Rev. **4**, 683–692 (2013)
4. Turney, P.D., Pantel, P.: From frequency to meaning: vector space models of semantics. J. Artif. Intell. Res., 159–164 (2010)
5. Park, W.-S., Kim, D.-K., Yang, Y.-K.: Driving characteristics classification of TCS data based on fuzzy c-means clustering algorithms. In: The 31st Conference of the Korea Information Processing Society (KIPS 2009), 23–24 April, pp. 1021–1024 (2009)
6. Bezdek, J.C., Ehrlich, R., Full, William: FCM: the fuzzy c-means clustering algorithm. Comput. Geosci. **10**(2–3), 191–203 (1984)
7. Yang, S., Choi, J., Bae, S., Chung, M.: A hybrid prediction model integrating FCM clustering algorithm with supervised learning. In: Park, D.-S., Chao, H.-C., Jeong, Y.-S., Park, J.J.(J.H.) (eds.) Advances in Computer Science and Ubiquitous Computing. LNEE, pp. 619–629. Springer, Singapore (2015)
8. The Kyunghyang Shinmun, http://www.khan.co.kr
9. Choi, K.-S., et al.: HanNanum, KAIST GPL v3 (1999)
10. Huh, J.H., Otgonchimeg, S., Seo, K.: Advanced metering infrastructure design and test bed experiment using intelligent agents. J. Supercomput. **72**(5), 1862–1877 (2016). Springer, USA

Information Reminder System Based on Word Registered by User

KyeYoung Kim[1], Byeong-Eon Ahn[1], Suk-Young Lim[1], Daejin Moon[2], and Dae-Soo Cho[1(✉)]

[1] Division of Computer Engineering, Dongseo University, 45, Jurae-ro, Sasang-gu, Busan, South Korea
kimlyl91@gmail.com, Lockon2200@gmail.com, djrwq@gmail.com, dscho@dongseo.ac.kr

[2] Doule-P Corporation, 47-207, Jurae-ro, Sasang-gu, Busan, South Korea
wizardyk@gmail.com

Abstract. According to the Hermann Ebbinghaus's forgetting curve, the memory of people starts forgetting after 10 min more than 50 % after one hour. The reason that people typically record words on note is for reminding the words. If people don't consciously search the recorded words on note, the words will be forgotten. This paper presents the information reminder system that helps a user to remind words. This system helps a user by giving the word highlighting the shades in case of a word that appear during web surfing.

Keywords: Forgetting curve · Reminder system · Word · Word highlighting

1 Introduction

Modern people have a high enthusiasm for learning various languages and terminology because of the global era advent. People were a lot of work to memorize the words.

People have researched various methods. Typical method for repeated studying is making a personal vocabulary list. However vocabulary list has various demerits. For example, if people don't consciously look up the recorded information on vocabulary list, the information will be forgotten. Also if people lose a personal vocabulary list, they lose all information. Personal vocabulary list depend on the memory of person to remind words. This mean that people should be always conscious what recorded information and vocabulary list. For this reason, Conventional method that makes a vocabulary list to remind words is not efficient method.

In this paper, we present a system that helps a user unconsciously reminding information. This system helps a user by giving the word highlighting the shades in case of a word that comes out during web surfing. Also this system offer trend information through web crawling for user.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_145

2 Literature Review

Hermann Ebbinghaus presented that memorization of word is necessary to repeat several times. However, when the iterative learning is also at a time interval of the excessively long period of time, it is not effective. That is, people have to repeat learning when the memory has forgotten, it is possible to enhance effectively the memory again, as shown in Fig. 1 [1].

Karpicke and Roediger presented that repeated retrieval during learning is the key to long-term retention. They emphasis that repeated test's learning effect is greater than repeated study [2].

Youm, Oh, and Chun presented English vocabulary learning application that based on forgetting curve. The application calculates learning time of words that user has learned and recommends a learning word that has fit learning time [3].

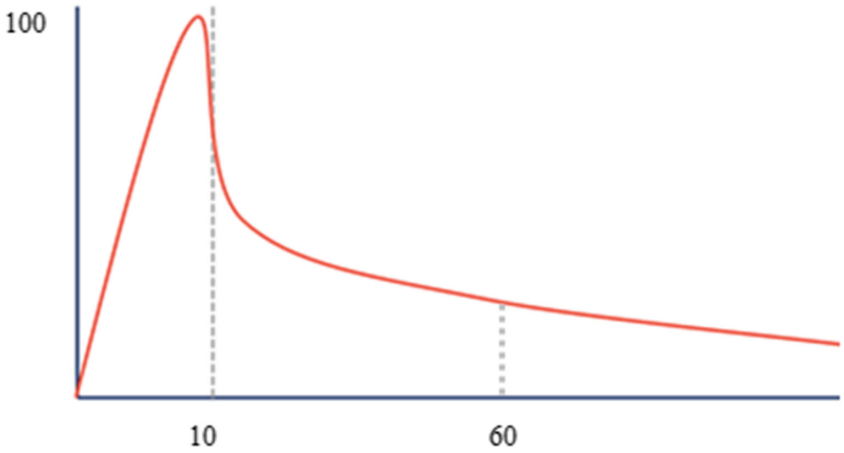


Fig. 1. Hermann Ebbinghaus's forgetting curve. This figure show that the memory of people start forgetting after 10 min more than 50 % after 1 h.

3 Methodology

3.1 Automatic Word Registration

Information reminder system supports automatic word registration. This system can recognize the word that you've been searching web dictionary. This system can automatically register a word and meaning of the word by reducing the hassles that you have to individually register the word.

3.2 Highlighting

When a user surfing the web, word that the user has registered appears in the web site, information reminder system highlight the word that put color shade, as shown in Fig. 2. The word that the user has registered can be highlighted by this way to the eye of the user. This function not only put the color shades but indicates the meaning of the collected words, when a user places the mouse over the word. In this ways, the system helps users to remind the words.

Fig. 2. Web browser applying Smart Reminder's highlighting function. This function can highlight the word that user register in web surfing.

3.3 Word Learning

When you generate for the first time of the words object, enter some information. Some information is composed of the word, information, maximum reminding stage, current reminding stage. The word is a collected word by user. Information is the word's meaning or data that user entered. Reminding stage is a representation of a degree of the user's words learned as a value.

The maximum reminding stage is user's the maximum target value to be learned, current reminding stage means the value of the degree learned so far. When you click on the word that became a highlight, it will be to adjust the transparency of the highlights to calculate the maximum reminding stage's value and the current reminding stage's value in percent. If the current reminding stage's value has reached the maximum reminding stage's value, the word means what has completed the learning to 100 % and the transparency of the highlights is terminated, as shown in Fig. 3.

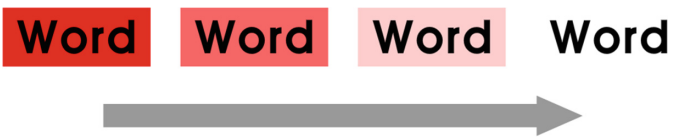


Fig. 3. Transparency stage of a highlighting word that user has registered.

3.4 Trend Information Collection

This system collects trend information including words registered by user from web articles.

First, it collects web articles by parsing their URL as interrogation of URL domain that using the words registered by user. Figure 4 is an example of HTML documents collected by web crawling. Second, it goes through the preprocessor phase for extracting main sentence from Web articles collected from HTML document type.
 Tag separating the paragraph in the preprocessor phase is not removed. However, unnecessary tags as , <script> are upper-removed, stored into an array in order after grasping the location of the start and end of a string, and only extract the text. This system separate the extracted main text to paragraph units through <br> tags and calculate frequency of words that user registered in each paragraph. Last, the most frequent paragraph among the paragraphs is mapped with the word registered by user. When the user is reminded of registered words and information by collected trend information, the user can obtain the information on the current issue.

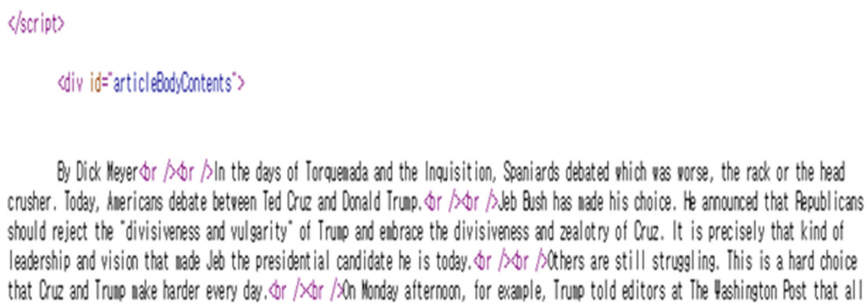


Fig. 4. This is html document example extracted by web crawling. Html document include html tags as , <script>.

4 Result

We implement information reminding system through word. We compare Smart Reminder with different learning method to evaluate against our system's performance during a week. The different learning method is formally iterative learning method that just writes a lot of words in note. In this Evaluation, we evaluate two groups. Group A use Smart Reminder. Group B use the different learning method to study a hundred words. We evaluate recall rates of a hundred words in two groups after a week.

Figure 5 reports the word's recall rates in two groups. We see that group A's recall rates less than group B's recall rates. However, Fig. 5 show that group A's growth of recall rates greater than group B's growth of recall rates. This result indicates what the system's learning effect close the iterative learning method in the long term.

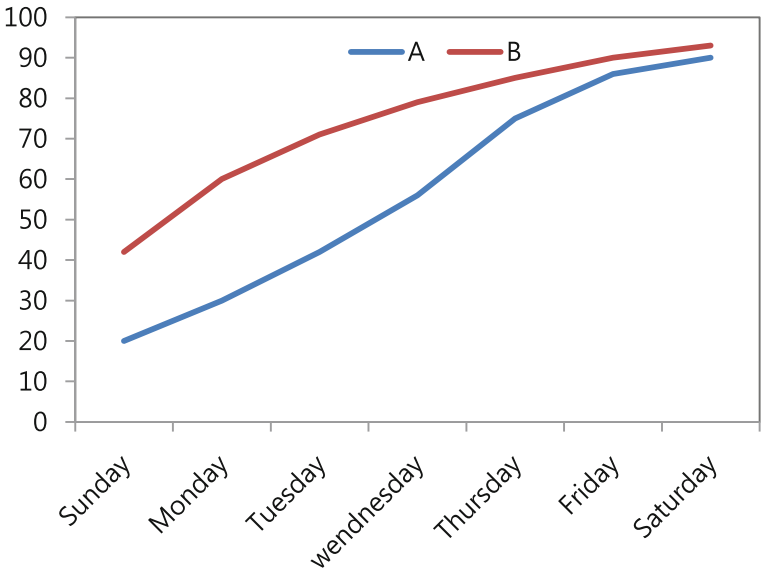


Fig. 5. This is recall rates of group A and B during a week. After a week, Group A and B's recall rates was similar.

5 Discussion and Conclusion

Information Reminder system has the strength that helps user to remind words on web surfing. However, this system's performance is less than iterative learning method. Yet, we cannot conclude that the system is not effective because the evaluation was done and had a little amount of words in the short term. We will research the method that the system's learning effect is greater than iterative learning method in the short term. We will improve the visual effect of a remind function than current system.

Acknowledgments. This research is results of a study on university for Creative Korea-1 (CK-1) Project and the Leaders in Industry-University Cooperation (LINC) Project, supported by the Ministry of Education.

References

1. Ebbinghaus, H.: Memory: a contribution to experimental psychology. Ann. Neurosci. **20**(4), 155–156 (2013)
2. Karpicke, J.D., Roediger, H.L.: Repeated retrieval during learning is the key to long-term retention. J. Memory Lang. **57**(2), 151–162 (2007)
3. Youm, K., Oh, K., Chun, Y.: English vocabulary learning application development applying forgetting curve and match result based rating system. J. Korea Game Soc. **15**(3), 151–160 (2015)

A Study of Determining Abnormal Behaviors by Using System for Preventing Agricultural Product Theft

Jin Su Kim[1], Min-Gu Kim[1], Byung Rae Cha[2], and Sung Bum Pan[3](✉)

[1] Department of Control and Instrumentation Engineering, Chosun University, Gwangju, Korea
babotn5@gmail.com, happy9433@gmail.com
[2] Super Computing and Collaboration Environment Technology Center, GIST, Gwangju, Korea
brcha@nm.gist.ac.kr
[3] Department of Electronics Engineering, Chosun University, Gwangju, Korea
sbpan@chosun.ac.kr

Abstract. At present, the environment of agricultural production faces reduced productivity due to population decline and aging population, and agricultural products may be stolen anytime. To solve such problems, researchers have studied IoT-based smart farms. A smart farm is a system for enabling users to manage and control agricultural crops anytime anywhere on the basis of IoT (Internet of Things). This study aims to examine a system for preventing agricultural product theft by using an image monitoring system, for example, CCTV, which is part of the smart farm 1st model functions. First, the system uses image information inputted by the image monitoring system to detect objects loitering agricultural crops, and then uses directional change information to determine objects loitering the crops as abnormal behaviors. The system then uses the single-board computer, Raspberry Pi, to output real-time warning sounds and send the information of the object determined as abnormal behaviors to the user to prevent agricultural product theft.

Keywords: Smart farm · Intelligent surveillance system · Abnormal behavior · Loitering detection · Raspberry Pi

1 Introduction

Presently, rural districts are gradually impoverished because of farm population decline and aging farmers resulting from young people going to urban areas to work and low birthrates. A survey reveals farm population was 3.43 million people in 2005 and is expected 2.34 million people in 2020. Aging farm population was 26.7 % in 2005 and 36.8 % in 2013 which is a great increase [1]. Criminal behaviors of agricultural product theft taking advantage of this point continue to occur. Agricultural product theft does damage to farmers in terms of mental health and economy [2].

To address this issue, smart farm systems have been studied. A smart farm system has functions of checking temperature in real time, remote control and CCTV security

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_146

monitoring. This study aims to examine a system for preventing agricultural product theft by using an image monitoring system, for example, CCTV. The adaptive difference image technique is used for the image inputted by the image monitoring system to detect moving objects, and morphology for enhancing the accuracy of object detection is applied to remove noise. When an object is detected, the balance points thereof is used to measure movement direction. A balance points depending on directional changes is given to determine abnormal behaviors by a person loitering crops. When it is determined as abnormal behaviors, Raspberry Pi is used to output real-time warning sound and send object information to a user to prevent agricultural product theft.

Section 2 of this study describes IoT system development used in the agricultural environment and issues involved therein as part of prior studies. Section 3 describes a smart farm and the system for preventing agricultural product theft which is a 1st model of the smart farm system. Section 4 describes conclusion and future studies for the system for preventing agricultural product theft.

2 Prior Studies

From 2004 to 2009, the R&D model project was conducted to develop IoT technology applicable to producing, distributing and consuming agri-food, supervised by the Ministry of Information and Communication, and the Ministry of Knowledge Economy. Exemplary technology used includes USN, RFID and LED, and GIS, GPS, QR code and bar code technology. From 2010 to 2013, the Ministry of Agriculture, Food and Rural Affairs developed models to spread the contents of the R&D model project conducted, and is spreading the IoT technology models (greenhouse farming, fruit tree growing, pig farming) of which outcome has been identified from 2014 to the present (Table 1).

Table 1. Smart agriculture related market present condition and prospect

Division	2012	2013	2014	2015	2016	CAGR (%)
Smart farm	13,378	14,274	15,231	16,251	17,340	6.7
Plant factory	500	767	1,175	1,800	2,759	53.3
Intelligent agricultural machine	10,417	12,500	15,000	18,000	21,600	20,0

Regardless of government's efforts for smart farming described above, farmers are not much satisfied with them. This may be based on diversified bodies and isolated promotion of projects and policies without establishing specific orientation, standard or direction [3].

At present, IoT technology and economic power of Korea is not so high in comparison with other countries, and key components (sensors, controllers, etc.) are imported from other countries to build a system. Therefore, because most farmers buy and install imported systems as a solution, prices, operation and maintenance costs are

so high, and lack of compatibility between components contributes to difficulty in maintenance.

To address this issue, this study uses Raspberry Pi which is relatively cheap and does not need specialized knowledge for operating image information of the image monitoring system to examine the system for preventing theft in real time.

3 IoT Based on Method for Detecting Abnormal Behavior

Figure 1 shows a smart farm 1st model suggested by the National Institute of Agricultural Sciences. The smart farm is based on IoT technology and a system for controlling and managing farms through wireless connection to the Internet. The smart farm 1st model has various functions of automatically opening and closing vinyl greenhouse doors, checking temperature in real time, remotely controlling the greenhouse, and monitoring with CCTV. This study is based on monitoring with CCTV which is part of functions of the smart farm 1st model to examine the system for preventing agricultural product theft.

Fig. 1. Example of smart farm 1st model based on IoT technology

Most behavior of criminals loiter the environment surrounding a place of their crime before actually conducting their criminal behaviors [4]. The literal meaning of loitering in this study is "roaming about places of interest here and there" [5]. Therefore, the person loitering a place has more changes in direction than those not loitering.

In this study, objects are detected by using the adaptive difference image technique, and morphology is conducted to enhance the accuracy of detection. To determine whether the detected object is loitering a place, that is, an abnormal behavior, the balance points of an object is used and a balance points depending on the amount of angular changes is given. When an abnormal behavior is determined, a warning sound is output and image information is sent to the system user in real time to prevent agricultural product theft.

3.1 Adaptive Difference Image

The difference image technique is used, which is a popular method used to detect objects in real time and does not require much computation relatively. However, one disadvantage of the difference image technique is wrong detection of movement resulting even from stop situations and changes in light after small movement of the

object over time because the background image is initially set up and not changed [6]. Furthermore, another issue is wrong detection due to natural phenomena occurring outdoors, for example, snowing or raining in the agricultural environment. The technique addressing this issue is the adaptive difference image. Adaptive difference image is a technique of giving a predetermined weight to the current image and the background image to update the background image where there is no movement. The method of detecting movement is conducted with the following Eq. 1. I_n is an image currently inputted; I_{n-1} is an image inputted previously; and Th_{moving} is a threshold for the movement.

$$I_n(x, y) - I_{n-1}(x, y) > Th_{moving} \tag{1}$$

Where the difference between the image currently inputted and the image inputted previously is greater than the threshold, the background image is not updated. Where it is smaller than the threshold, the area is determined as an area without movement to update the background image with Eq. 2. $B_n(x, y)$ is the current background image; $B_{n+1}(x, y)$ is the next background image; and α is a weight.

$$B_{n+1}(x, y) = \begin{cases} B_n(x, y) & \textit{Moving Detection} \\ \alpha B_n(x, y) + (1 - \alpha) I_n(x, y) & \textit{Not Moving Detection} \end{cases}. \tag{2}$$

Figure 2 shows how to detect an object by using adaptive difference image in an indoor environment. Figure 2(a) shows an original image with no moving object detected, and Fig. 2(b) shows a resulting image after conducting adaptive difference image technique for Fig. 2(a). Figure 2(c) shows an original image where a moving object is detected, and Fig. 2(d) shows a resulting image after conducting adaptive difference image technique for Fig. 2(c).

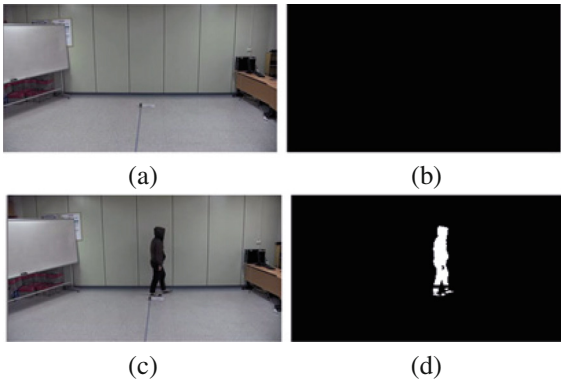


Fig. 2. Object detection in video using adaptive difference image, (a) Non object image, (b) Adaptive difference image of non-object, (c) object image, (d) adaptive difference image of object

3.2 Measuring Angular Changes Between Exemplary Balance Points

The method of using angular changes between exemplary balance points is used to detect persons loitering a place of interest. The balance point of an object at given time intervals is an exemplary balance point. Figure 3 shows the method of obtaining an angle between the exemplary balance points. The angle θ between a straight line 1 and a straight line 2 is obtained by drawing the straight line 1 connecting C_{t-1} to C_t, and the straight line 2 horizontally connected to C_{t-1}. After, measure angular changes and obtain a weight depending on the changes.

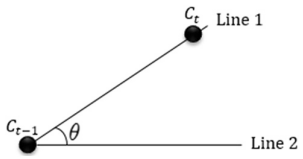


Fig. 3. One measuring the angle between the balance points, C_{t-1} is prior balance point, C_t is present balance point, angle θ

Figure 4 shows the standard of angular changes. The part the arrow indicates is the direction of current object movement, and the straight line with an indication of 0° is Line 2 shown in Fig. 3. Table 2 illustrates weights depending on angular changes. A greater weight is given to greater object direction changes, but a smaller weight to smaller object direction changes.

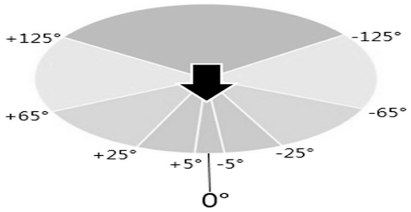


Fig. 4. Change in the angle between balance points

Table 2. Weighted according to the angle change.

Angle change	Weight
−5 ∼ 0 or 0 ∼ 5	0.01
−25 ∼ −5 or 5 ∼ 25	0.02
−65 ∼ −25 or 25 ∼ 65	0.04
−125 ∼ −65 or 65 ∼ 125	0.07
Under −125 or Over 125	0.10

4 Conclusion and Future Work

Raspberry Pi used for automatically sending e-mails, SSMTP was used to make an SMTP socket, and ALSA was used to use an audio port to output warning sounds. Figure 5(a) shows detection of an object loitering an indoor environment at a distance of 5 m between the object and the camera. Figure 5(b) shows detection of an object loitering an outdoor environment at a distance equal to or farther than 20 m between the object and the camera. Figure 5(c) shows a movement trajectory by using exemplary balance points of an object, and Fig. 5(d) shows sending object information to the user where the movement is determined as an abnormal behavior.

This study aims to examine a system for preventing agricultural product theft by monitoring with CCTV which is part of functions of the smart farm 1st model. Abnormal behaviors are determined after detecting an object of abnormal behaviors to prevent theft. It was identified that a warning sound is output in real time, and the object image information is sent to the user after determining a person of abnormal behaviors. Therefore, the system may prevent agricultural product theft in advance, and provide a crucial evidence to arrest criminals by sending image information of the loitering site. A future plan is to diversify a database to apply the system to various environments, and use other learning methods than angular information to improve loitering detection performance.

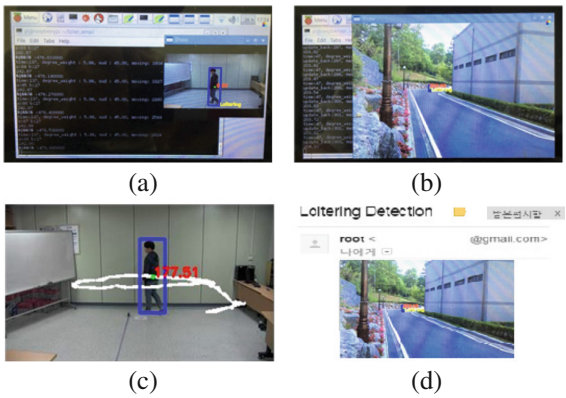


Fig. 5. Experimental result, (a) Short distance loitering detection, (b) Long distance loitering detection, (c) Object movement trajectory, (d) Automatic sending e-mail

Acknowledgments. This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (2015H1C1A1035823).

References

1. Social Indicators easy to understand Commentary. Statistics Korea (2011)
2. Kim, D.J., Lee, S.H.: A security system of agricultural products based on read time event recognition. Korean Inst. Intell. Syst. **22**(1) (2012)
3. Kim, G.J., Huh, J.D.: Trends and prospects of smart farm technology. Electron. Commun. Trend Anal. (2015)
4. Kang, J.H., Kwak, S.Y.: Loitering sudden running and intruder detection for intelligent surveillance system. Korea Inf. Sci. Soc. **39**(1), 353–355 (2012)
5. Chang, I.S., Cha, H.H., Park, G.M., Lee, K.J., Kim, S.K., Cha, J.S.: A study of scenario and trends in intelligent surveillance camera. Korea Inst. Intell. Transp. Syst. **8**(4), 93–101 (2009)
6. Lee, C.S., Jun, M.S.: Motion detection using adaptive background image and a net model pixel space of boundary detection. J. Korean Inst. Commun. Sci. **30**(3), 92–101 (2005)

A Study of Simple Classification of Malware Based on the Dynamic API Call Counts

Jihun Kim[1], Seungwon Lee[1], Jonghee M. Youn[1(✉)], and Haechul Choi[2]

[1] Department of Computer Engineering, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea
{fl3521, ll02lsw}@naver.com, youn@yu.ac.kr
[2] Department of Multimedia Engineering, Hanbat National University, Daejeon 34158, South Korea
choihc@hanbat.ac.kr

Abstract. Recently, as the rapid development of the Internet enabled easy downloading of diverse files, the number of cases of file download from unreliable paths has been increasing. This situation is advantageous in that accessibility to information is improved while being disadvantageous in that there is no defense against exposure to malware. The present paper proposes a method of judging whether programs are malicious based on Cuckoo Sandbox, which is a dynamic malware analysis system and classify the programs by comparing malware programs collected and classified in advance based on the dynamic API call counts of the programs.

Keywords: Malware · Classification · Cuckoo Sandbox · API call count

1 Introduction

Currently, the kinds and numbers of malware are rapidly increasing. In the times when personal information such as financial information is processed through the Internet, increases in malware indicate dangerous situations. In addition, as the number of cases of downloading diverse kinds of files through the Internet, the number of downloading of unreliable files has also increased. Downloading through unreliable paths are highly likely to lead to receiving programs infected with malware such as viruses and the Trojan horse so that the downloading user's information is leaked unnoticed and the malware is propagated to other computers to cause secondary and third damage of information leakage.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1A2B1012652, the MSIP(Ministry of Science, ICT and Future Planning, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-R2718-16-0035) supervised by the IITP (National IT Industry Promotion Agency), the Basic Science Research Program through the NRF funded by the Ministry of Education (NRF-2015R1C1A1A02037561) and the 2016 Yeungnam University Research Grant.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_147

To respond to the foregoing problems, many studies are in progress to find ways to distinguish diverse kinds of malware. The present paper proposes a method of analyzing programs to classify the programs into different kinds of malware based on the dynamic API Call counts of the codes of the programs.

The present paper is composed as follows. In Sect. 2, malware analysis methods using API and dynamic analysis related studies are introduced and in Sect. 3, background knowledge of API and Cuckoo Sandbox, which is the basis of the present study is explained. In Sect. 4.2, API extraction using manufacturing tools, tools used in malware analysis such as similarity digitization, and malware classification methods are presented and in Sect. 4, related experiments and the results are explained. Finally, in Sect. 5, conclusions and the direction of future studies are presented.

2 Related Works

Since the risk of malware has been judged to be high, there are many results of previous studies on malware analysis methods conducted to prevent malware. The core of the present paper is dynamic analysis methods and malware analysis methods using APIs.

Malware analysis methods are largely divided into static analyses and dynamic analyses. Among them, in the present paper, a dynamic analysis method is used that constructs a certain analysis environment on a virtual machine, executes the malware, and analyzes the behavior of the malware [1].

Among dynamic analysis methods, a method that would create detection patterns using the bypassing technique of executable compression to detect malware was proposed [2]. The proposed method creates signatures through a process of integration of the behavior and sequential patterns of malware using the API of the malware. Another proposed method extracts functions called in the system and links the functions with the degree of Native API Call hiding occurring in kernel units to depend against and prevent malware programs that are not standardized. This method is used as a technique to enhance the efficiency of Call based detection [3].

API based malware analysis techniques require the results of APIs extracted from malware samples. First, the new malware is analyzed using the results of extracted malware APIs. Then, the similarity of the new malware is compared with that of existing malware programs to classify the new malware [4].

Cuckoo Sandbox [5] is an open source project that support users for convenient identification of malware by analyzing the complex behavior of malware executed on virtual machines. Cuckoo Sandbox is used to diagnose malware by using a guest PC-virtual machine(basically virtual machine) to analyze files that are suspected as malicious files in the virtual machine and store only the resultant values so that malicious files can be safely analyzed.

3 API Call Count Based Malware Classification Method

For a program to interact with other programs or use system resources, the necessary APIs should be called and delivered to the operating system. Which APIs are called and used becomes an important characteristic in defining the behavior of the program.

Therefore, in the present paper, to identify the characteristics of individual programs, the frequencies of calls of APIs are investigated and the results are used in similarity analysis.

Utilizing Cuckoo Sandbox, the data on malware collected in advance are extracted through dynamic analysis by hooking APIs being called. All the APIs called by the malware are extracted and aligned based on the counts of calls. The extracted APIs are classified again according to the purposes of attacks. Among the APIs gathered by the purpose of attacks, 10 APIs with the highest call frequencies in each folder are designated as a set to define the set as the characteristics of the relevant malware. In addition, the characteristics are used as criteria for judgment whether files are malicious. However, the attacking methods of individual malware programs cannot be classified by just gathering APIs with high call counts become some APIs are commonly called frequently.

Therefore, in the present paper, a method that can specify those APIs that appear frequently only in those malware programs that belong to certain classes is proposed. Malware programs of which the class is already known is defined as a group and the APIs and their counts called by individual malware programs belonging to the class are extracted. We defined nine malware classes and underwent the foregoing process for 20 malware samples in each class. The following algorithm was designed to specify those APIs that are frequently called only by certain classes of malware programs but not by other classes based on the API call information extracted from the malware programs class as such.

First, add up all API Call counts extracted from malware programs belonging to the same class to integrate API call counts by class. Thereafter, compare and analyze the integrate API call counts by malware class. That is, compare the integrate API call counts of one class with those of the remaining eight classes repeatedly by deducting the average call count for an API of the remaining eight classes from the call count for the same API of one class. After repeating the comparison as such, those APIs that are characteristic in individual classes should maintain high values and those APIs are frequently called by all classes should have small values.

Tables 1 and 2 show top 10 APIs with high frequencies in backdoor malware and spy malware respectively.

After securing characteristic APIs in individual malware classes as such, classify those malware programs that have not been classified yet based on the foregoing results. Obtain the dynamic API Call Sequences of individual codes or programs that must be judged whether malicious or not using Cuckcoo Sandbox. Obtain similarity values by comparing the entire API counts with the total counts of those APIs in the API Sets by class.

※Value of similarity to each malware class $X_M = \sum (M_{Set} \cap I_{API^{10}})$

M = Malicious class(ex. Dialer, Dropper)

I = Suspected file to be analyzed

Set = API Set

API^{10} = Top 10 APIs in the API list

Table 1. Backdoor set

NtDeviceIoControlFile
VirtualProtectEx
WriteConsoleW
RegSetValueExA
RegCreateKeyExA
DeviceIoControl
bind
socket
connect
closesocket

Table 2. Spyware set

ReadProcessMemory
NtQueryDirectoryFile
NtOpenFile
NtQueryInformationFile
LdrGetProcedureAddress
RegCloseKey
RegQueryValueExW
ZwMapViewOfSection
NtWriteFile
RegOpenKeyExW

The similarity values for individual classes are obtained as such and the class for which the API Call Sequence has the largest value when seen in percentiles becomes the final class of the API Call Sequence.

4 Experiment and Results

To experiment the method proposed in the present paper, malware programs were classified into nine types. Approximately 20 malware samples of each of Backdoor, Dialer, Downloader, Dropper, Injector, Rootkit, Spyware, Trojan, and Worm were collected. API lists were extracted using the collected malware samples and API Sets were designated.

4.1 Results of Malware Classification of New Programs

Using the logical formula and API Sets mentioned above, a malignancy judgement tool using actual API Call Sequences was made and the results of the tool after being used on programs suspected of Dialer attacks, Spyware attacks, or Injector attacks were output (Tables 3 and 4).

As can be seen in the above tables, if the method proposed in the present paper is applied to actual suspected programs, the program will be diagnosed for its purpose of attacks with digitized values. However, since the values of other purposes of attacks obtained as such are similar in many cases, (Spyware continuously recorded high values) files may be wrongly diagnosed although the probability is low.

In the case of Table 5 shown above, it can be seen that the relevant program has characteristics of not only Injector malware but also of many other malware programs. That is, programs may have the characteristics of multiple malware programs instead of having the characteristics of only one malware program. In such cases, if several malware programs are selected and the similarity levels are compared, the accuracy of the result should be higher.

Table 3. Suspect of dialer attacks

Similarity values		Percentile
Backdoor	12	0.37
Dialer	760	23.67
Downloader	13	0.41
Dropper	683	21.27
Injector	148	4.61
Rootkit	288	8.97
Spyware	710	22.11
Trojan	42	1.31
Worm	555	17.28
Total	3211	100

Table 4. Suspect of spyware attacks

Similarity values		Percentile
Backdoor	43	2.83
Dialer	235	15.48
Downloader	23	1.51
Dropper	255	16.79
Injector	283	18.64
Rootkit	28	1.84
Spyware	469	30.89
Trojan	68	4.48
Worm	114	7.51
Total	1518	100

4.2 Results of Malware Classification of APIs in Existing Folder

In addition to the codes of new programs, experiments were conducted to see whether the APIs currently classified by malware folder are actually highly similar to the relevant malware. Five APIs were randomly extracted from each malware folder. The results obtained from the extracted malware programs using the distinguishing tool are as shown in the following Tables.

Table 5. Suspect of injector attacks

Similarity values		Percentile
Backdoor	0	0
Dialer	118	4.34
Downloader	45	1.66
Dropper	112	4.12
Injector	1112	40.94
Rootkit	31	1.14
Spyware	1085	39.95
Trojan	53	1.95
Worm	160	5.90
Total	2716	100

Table 6. List of APIs

Dropper folder	Injector folder	Dialer folder
1. Dropper 24.2 % (1)	1. Injector 40.1 % (1)	1. Dialer 19.1 % (2)
2. Dropper 21.8 % (1)	2. Injector 48.7 % (1)	2. Dialer 19.4 % (2)
3. Dropper 23.1 % (2)	3. Injector 40.9 % (1)	3. Dialer 23.6 % (1)
4. Dropper 28.1 % (1)	4. Injector 66.7 % (1)	4. Dialer 32.4 % (1)
5. Dropper 16.2 % (2)	5. Injector 94.5 % (1)	5. Dialer 18.4 % (2)

As can be seen in Table 6, APIs in the relevant folders show high similarity to individual folders. It can be seen that the relevant malware programs are included at least top two in call counts. This supports the results indicating that the similarity algorithm is valid.

5 Conclusion and Future Works

In the present paper, a method was proposed to judge whether malware exist in programs and determine the types of malware. For the proposed method, all APIs of sample malware programs were extracted and collected. Based on the collected APIs, when a new program or code came in, similarity levels to the malware programs are calculated. This way, whether the program or code is similar to any malware type or whether it is malicious can be judged.

As for the direction of future studies, since the proposed method is based on malware data collected in advance, to detect more diversified malware, more types of sample malware will be collected to enhance the accuracy.

References

1. Han, K.-S., Kim, I.-K., Im, E.-G.: Malware family classification method using API sequential characteristic. J. Secur. Eng. **8**(2), 319–335 (2011)
2. Park, N.-Y., Kim, Y.-M., Noh, B.-N.: A behavior based detection for malicious code using obfuscation technique. J. Korea Inst. Inf. Secur. Cryptology, June 2006
3. Kang, T.-W., Cho, J.I., Chung, M.-H., Moon, J.-S.: Malware detection via hybrid analysis for API calls. J. Korea Inst. Inf. Secur. Cryptology, December 2007
4. Park, J.-W., Moon, S.-T., Son, G.-W., Kim, I.-K., Han, K.-S., Im, E.-G., Kim, I.-G.: An automatic malware classification system using string list and API. J. Secur. Eng. **8**(5), 611 (2011)
5. Cuckoo Sandbox. http://www.cuckoosandbox.com

A Low-Power Sensing Management Method for Sustainable Context-Awareness in Exclusive Contexts

Dusan Baek[1], Jae-Hyeon Park[1], Byungjeong Lee[2], and Jung-Won Lee[1(✉)]

[1] Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea
whitedusan@gmail.com, jaehp2560@gmail.com, jungwony@ajou.ac.kr

[2] Department of Computer Science and Engineering, The University of Seoul, Seoul, South Korea
bjlee@uos.ac.kr

Abstract. A context-aware service makes the assumption that two or more exclusive contexts are not inferred simultaneously such as driving and studying, indoor and outdoor. However, in practice they are sometimes inferred at the same time, and it causes inefficient power consumption because it does not make sense. To handle this problem and improve power efficiency, we propose a low-power sensing management method for sustainable context-awareness in exclusive contexts. In our method, we identify the exclusive contexts by using sensing models. Then, we determine next sensing time by utilizing supplementary sensor or increase the period of sensing in the exclusive contexts (i.e., back off). The results of our preliminary application show that the power efficiency is improved to 21 % in the exclusive contexts. The proposed method will be more effective when the exclusive contexts are inferred more frequently according to diffusion of context-aware services in the future.

Keywords: Low-power · Context-aware · Sustainable · Sensor manage

1 Introduction

The diffusion of the mobile device equipped with various sensors makes it possible for users to utilize context-aware service. The context-aware service senses physical environment around the user by using the various sensors to offer well-fitted service. However, the context-aware service requires sensing continuously for monitoring user's context, so it consumes much power. Especially, it becomes more important issue for the mobile device that is supplied power from limited battery [1].

For this reason, there have been a large number of studies for sustainable context-awareness. They try to identify causes of the inefficient power consumption or remove the causes. However, the existing studies have been lack of careful consideration of the environment in which the heterogeneous context-aware services exist. Particularly, they do not consider the exclusive contexts, which cloud lead unnecessary

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_148

power consumption. For example, in indoor/outdoor context-aware service, indoor and outdoor contexts could be inferred when the Wi-Fi signature is acquired, and the GPS signal is received, respectively. Although indoor and outdoor contexts cannot coexist logically at the same time, they are sometimes inferred by Wi-Fi and GPS sensors at once like at an edge of the building. Similar to this case, inferring two or more exclusive contexts simultaneously is meaningless semantically and leads inefficient power consumption, so we have to handle the problem for sustainable context-awareness.

In this paper, to solve the problem, we propose a low-power sensing management method. In our method, we define each context occured in the context-aware service as a sensing model. Then, we merge them into integrated sensing model with parallel composition, and identify the unnecessary sensing states inferring the exclusive contexts. Finally, we determine next sensing time by utilizing supplementary sensor or increase the period of sensing in exclusive contexts. We conduct preliminary application of our method, and the results show that the power efficiency is improved to 21 % in the exclusive contexts.

2 Related Work

There have been many studies for sustainable context-awareness. They could be classified into the studies for identifying causes of inefficient power consumption and for removing the causes. First, the goal of the former studies is to help researchers or developers localize the energy bug (also called energy black hole or energy leak) that causes the inefficient power consumption [2–4]. Each of them makes power model at the component level [2, 3] or at the instruction level [4], and they estimate and analyze the power consumption of context-aware service by using it. However, they have shortcoming that only provide the information about amount of the power consumption without the additional information. To remedy the shortcoming of previous works, [5] proposes an approach to diagnose energy inefficiency problems systematically. The approach provides the information about misusage of the sensor listener, and underutilization of the sensory data by analyzing the sensor utilization. These studies for identifying causes of inefficient power consumption prepare the ground for the low-power-related study for sustainable context-awareness. In this paper, by using the methods and results of previous studies, we obtain the information about the sensing state inferring the exclusive contexts, evaluate our method.

Second, the studies for removing the causes of inefficient power consumption are mainstream in this domain because reducing power consumption in the mobile device is critical issue. [7] presents APE, an annotation language and middleware service that eases the development of energy-efficient Android application. [1] presents rate-adaptive positioning system that adjusts the periodically duty-cycle of GPS according to required accuracy. [8] presents a design framework for sensor management to substitute existing sensors to minimum set of sensors that consume lower power. Finally, [9] propose ACE (Acquisitional Context Engine) that is a middleware exploits sensing data cache or infer a context form another already-know context without redundant additional sensing. These studies provide the solutions for addressing the

inefficient power consumption, but they do not consider the exclusive contexts. Therefore, it is hard to consume the power efficiently in the exclusive contexts.

Our study aims to promote the power efficiency by managing the unnecessary sensing activity in the exclusive contexts. For this, we analyze the power consumption by using the tool of [6] that does not require the power model. Moreover, we utilize the sensing model based on automata that is commonly used by event-driven approaches [8, 11]. In order to manage the sensors, we adjust the period of sensing or determine next sensing time by employing the middleware like previous studies does.

3 Context-Awareness Considering Exclusive Contexts

Our method proposed in this paper exploits the impossibility of coexistence in exclusive contexts. To do this, we point out the unnecessary sensing state inferring the two or more exclusive contexts in the integrated sensing model, and manage the sensors to operate better efficiently. In this section, we state problem definition before explanation how to manage the sensors in the exclusive contexts.

3.1 Problem Definition

This subsection presents the problem definition to help understanding by describing a scenario. The scenario is as follows.

Scenario. Tom, a graduate student, usually takes the car to the lab, studies in the lab and attends the seminar with a professor. He utilizes context-aware service that provides three key functions. The service blocks a call during driving and sends the caller the message to inform his driving state. In addition, the service changes the ringer on vibrating mode during studying and mute mode during attending the seminar. The service judges three kind of context as follows: (1) The driving context if the GPS sensor receives the satellite data. (2) The studying context if the Wi-Fi signal of the office/laboratory can be received and sound around the device is greater than the pre-defined threshold. (3) The seminar context if the Wi-Fi signal of the office/laboratory can be received and sound around the device is less than the pre-defined threshold.

The driving, studying and seminar context are the exclusive contexts that cannot coexist logically at the same time. However, if the GPS sensor receives the satellite data and the Wi-Fi sensor can receive the Wi-Fi signal of the office/laboratory, then the driving and studying context or the driving and seminar context will be inferred. That two or more exclusive contexts are inferred simultaneously is meaningless semantically, so the sensing activity inferring these contexts is unnecessary. Therefore, sensing to infer these kinds of meaningless contexts has to be managed for sustainable context aware.

3.2 Low Power Sensing Management Method in Exclusive Contexts

Figure 1 describes the procedure of proposed method in this paper. First, we define each sensing model for context-awareness as automata, and merge the sensing models into integrated sensing model. Then, we identify meaningless sensing state, where infer the exclusive contexts, in the model. Finally, by employing the middleware, we determine next sensing time by utilizing supplementary sensor or increase the period of sensing in the exclusive contexts.

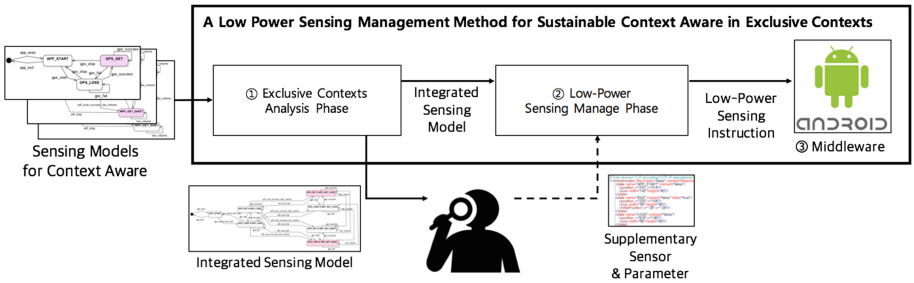


Fig. 1. The procedure of the low-power sensing management method for sustainable context-awareness in exclusive contexts

① *Exclusive Contexts Analysis Phase*

In exclusive contexts analysis phase, we merge sensing models as sources into the integrated sensing model as outcomes, and then analyze the integrated sensing model to identify the exclusive contexts. For this, we have to pre-define all the sensing models inferring the context as automata. Among the states in the model, the state inferring the context should be marked, and it is used for identifying the exclusive contexts. Additionally, in the model, the events and states provide the information to control the sensing activity in middleware. After constructing the sensing model, we merge them into the integrated sensing model through parallel composition. The integrated sensing model describes the whole sensing mechanism of the context-awareness, and the marked states indicate the sensing states inferring the exclusive contexts.

② *Low-Power Sensing Manage Phase*

Two or more exclusive contexts inferred at the same time are meaningless contexts. Therefore, we should manage the sensing activity in the sensing states inferring the exclusive context for low-power sensing. To do this, we use two methods for low-power sensing management.

First, we adapt the linear back off algorithm with the maximum value to the period of sensing in exclusive contexts. It improves power efficiency by increasing the period of sensing when the unnecessary sensing activity is detected (e.g., the periods of sensing are that 3 s, 6 s, 9 s, 9 s, 9 s, …). This method is easy to implement, and does not require to control another sensor. However, it has the critical disadvantage that detecting the transition from the exclusive contexts to normal context can be delayed.

The second method is using the supplementary sensor to determine the next sensing time for avoiding the unnecessary sensing activity. For example, when the sensing data of both GPS and Wi-Fi sensor infer the exclusive contexts, we can assume that if there is no movement, the exclusive context does not make the transition. Thus, we can decrease the power consumption by delaying the operation of the GPS and Wi-Fi sensor until the movement is detected via the 3-axis sensor that consumes less power than them. This method retains the reactivity in spite of spending low-power, but we must know the type and parameters of the supplementary sensor.

③ *Middleware for Sustainable Context-Awareness*

The middleware operates the sensors according to the low-power sensing instruction generated in previous phase. Hence, the middleware should have the ability to control the various sensors that perform the instruction. In our preliminary application, we implement the middleware for the GPS, Wi-Fi and microphone sensor in Android.

4 Application

We apply our method to the *Nexus 4* device for evaluating the applicability and the feasibility of it. The scenario used in the preliminary application is as follows. First, we maintain the context which the context-aware service infers two exclusive contexts simultaneously by using the microphone, Wi-Fi and GPS sensors. Second, after about 179 s (the vertical line), we transfer our context to normal context which the context-aware service infers only one context at once. Figure 2 shows the results of our preliminary application.

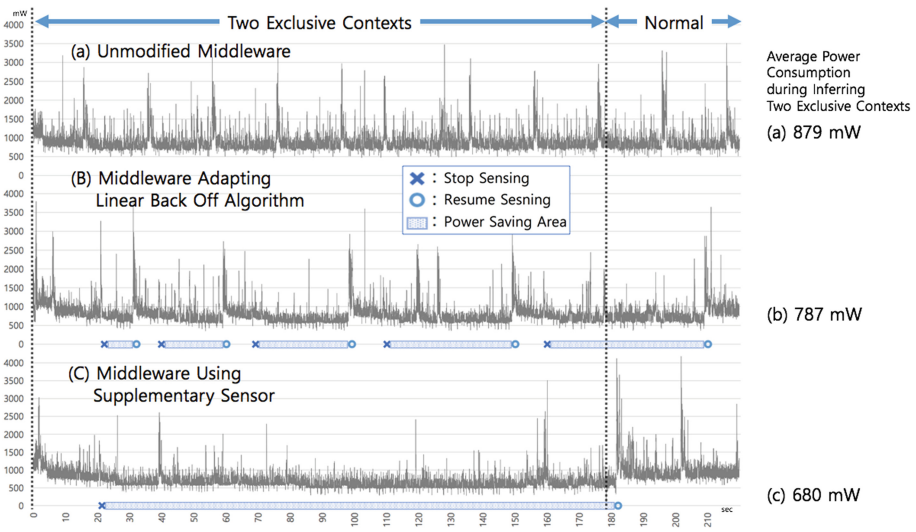


Fig. 2. The power consumption of a context-aware service in the transition from inferring two exclusive contexts to inferring one context.

As Fig. 2(a), the existing middleware cannot react in two or more exclusive contexts, so it leads the inefficient power consumption. On the other hand, as shown in Fig. 2(b) and (C), the middleware proposed in this paper can manage the sensor for efficient power consumption by increasing the period of sensing or by determining next sensing time. In the case of Fig. 2(b), we adapt the linear back off algorithm into middleware (10 s, 20 s, 30 s, …, max = 50 s). The result shows that the power efficiency is improved to 12 % compared to the unmodified middleware, but the response time to the transition is postponed because of the increasing the period of sensing. In the case of Fig. 2(c), we utilize the 3-axis sensor as the supplementary sensor to determine the transition. The result shows that the power efficiency is improved to 29 % compared to the unmodified middleware evaluated over the above scenario, also the response time is as fast as the unmodified middleware.

5 Conclusion

In this paper, we proposed the low-power sensing management method for decreasing the inefficient power consumption occurred when the sensors infer the exclusive contexts. In our method, we define the sensing models for context-awareness and merge them into the integrated sensing model. Then, we identify the unnecessary sensing states inferring the exclusive contexts, and control the sensing activity in the states by increasing the period of sensing or by determining next sensing time with supplementary sensor. The result of preliminary application shows 21 % improvement of power efficiency on average over our scenario. For the future work, we will create the mapping library between the primary sensors and the supplementary sensors and implement the middleware to enable control of all sensors.

Acknowledgments. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030504) and also supported by the National Research Foundation of Korea (NRF) grant funded by the MSIP (NRF-2016R1A2B1014376).

References

1. Paek, J., Kim, J., Govindan, R.: Energy-efficient rate-adaptive GPS-based positioning for smartphones. In: 8th ACM Mobile System, Applications, and Services, pp. 299–314. ACM, June 2010
2. Pathak, A., Hu, Y.C., Zhang, M.: Where is the energy spent inside my app? Fine grained energy accounting on smartphones with Eprof. In: 7th ACM European Conference on Computer Systems, pp. 29–42. ACM, April 2012
3. Hao, S., Li, D., Halfond, W.G., Govindan, R.: Estimating mobile application energy consumption using program analysis. In: 35th International Conference on Software Engineering, pp. 92–101. IEEE, May 2013

4. Lee, S., Jung, W., Chon, Y., Cha, H.: EnTrack: a system facility for analyzing energy consumption of Android system services. In: International Joint Conference on Pervasive and Ubiquitous Computing, pp. 191–202. ACM, September 2015
5. Liu, Y., Xu, C., Cheung, S.C.: Where has my battery gone? Finding sensor related energy black holes in smartphone applications. In: Pervasive Computing and Communications, pp. 2–10. IEEE, March 2013
6. Choi, K.-Y., Lee, J.-W.: Portable power measurement system for mobile devices. J. KIISE Comput. Practices Lett. **20**(3), 131–142 (2014)
7. Nikzad, N., Chipara, O., Griswold, W.G.: APE: an annotation language and middleware for energy-efficient mobile application development. In: 36th International Conference on Software Engineering, pp. 515–526. ACM, May 2014
8. Wang, Y., Lin, J., Annavaram, M., Jacobson, Q.A., Hong, J., Krishnamachari, B., Sadeh, N.: A framework of energy efficient mobile sensing for automatic user state recognition. In: 7th ACM Mobile System, Applications, and Services, pp. 179–192. ACM, June 2009
9. Nath, S.: ACE: exploiting correlation for energy-efficient and continuous context sensing. In: 10th ACM Mobile System, Applications, and Services, pp. 29–42. ACM, June 2012
10. Shih, E., Bahl, P., Sinclair, M.J.: Wake on wireless: an event driven energy saving strategy for battery operated devices. In: 8th Annual International Conference on Mobile Computing and Networking, pp. 160–171. ACM, September 2002

Content-Based Conformance Assurance Between Software Research Documentation and Design Guideline

Jong-Hwan Shin[1], Du-San Baek[1], Byungjeong Lee[2], and Jung-Won Lee[1(✉)]

[1] Department of Electrical and Computer Engineering, Ajou University, Suwon, Republic of Korea
{sjhl334, jungwony}@ajou.ac.kr, whitedusan@gmail.com
[2] Department of Computer Science and Engineering, The University of Seoul, Seoul, Republic of Korea
bjlee@uos.ac.kr

Abstract. Research-oriented software groups are groups that carry out research on original technology for software. The groups on development phase experience poor documentation because of two reasons. One is the lack of resources (i.e. time, costs) since the development phase is much shorter than their research phase. The other one is that the artifacts they worked on research phase are rarely used on the development documents. Therefore, we propose a method that can reduce poor documentation regarding their research documents and development (R&D) documents. We construct design guidelines from best practices and represent it by queries of semantics-aware traceability links. Then, we use a semi-automated method of conformance assurance between R&D documents with guidelines. Finally, we provide an explanatory guideline to assessment results. We evaluated documents generated from our previous R&D project to show the possibility of our method. Our method can help software R&D project documents for better quality with reduced time.

Keywords: Conformance assurance · Relevance Link Information Model (RLIM) · Traceability · Documentation quality

1 Introduction

Research-oriented software group is a group to study original technology for software. These groups need to develop the software to demonstrate their research. However, their developments experience poor documentation. Because their development is done in much less time than the research phase, therefore they have very little resources in time, money and manpower for documentation [8]. Moreover, they can't directly use their research artifacts when they writing development documents since research artifacts are composed of highly-abstracted contents while development documents are composed of detailed requirements [4].

With poor documentation, they face 'technical debt'. Technical Debt is a term which describes 'the extra development work acquired when engineers take shortcuts

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_149

that fall short of best practices [2]'. Although not all technical debt is bad, technical debt grows along development because it has interests just like financial debt. Therefore, unpaid technical debt may be presented as delay of project from resource-consuming tasks like fixing complicated bugs and fault localization. Many existing studies address poor documentation since technical debt from poor documentation is frequent. However, they focus on requirements traceability of development documents and they don't consider semantics for high-level qualities such as conformance between two items. Their approaches cannot improve research artifacts because the research artifacts do not usually have requirements for development.

In this paper, we propose a content-based method for assurance of the quality of research documentation including development documents to address problems described above. First, we define design guidelines, which reflect best practices of software development with consideration of research artifacts. A design guideline is made of a goal model and an explanatory guideline explains the goal model for user feedback. Then, we transform the design guidelines as queries of semantic-aware traceability links for automatic evaluation. We used an expert system to automatically evaluate documents with the transformed guidelines. The result of evaluation is given as form of a report that user can easily understand.

2 Related Work

Since poor documentation is one of major reason of technical debt [2], there have been many studies for improving software documents. First, there are automated approaches for analysis and evaluation of document quality [7, 9, 12]. Second, there are some approaches that help minimize time-consuming tasks for reviewers on review process [3, 13].

Wilson et al. [13] perform keyword-based analysis and quality measurements of software requirements. They set quality related terms and measured how frequently those terms are presented. This work shows the need for quality control of software documents and even a simple method can improve document quality, but this can't evaluate high-level quality like traceability directly. Jain et al. [9] uses controlled natural language approach for requirements analysis. The method performs lexical analysis for conformance of template. Then, it performs semantics analysis for completeness with state machines of each requirement. The system also generates a helpful message for unfinished specifications. However, this method only focuses the syntax of requirements. Dautovic et al. [7] uses visitor pattern to traverse document contents and simple rule-checking mechanism for quality measurement. However, the rules they created focus on the structural and format therefore semantics cannot be investigated.

Shen et al. [12] represented traceability of documentation contents with simple linked list instead of complicated graph. They gave the traceability linked list to regulator for helping regulatory review process. However, they didn't consider the evaluation of document quality and didn't considered explanatory guidelines which is useful for developers. Antonino et al. [3] they suggest parameterized safety requirement templates to ensuring traceability throughout software documentations in safety-critical system domain. They used controlled natural language to avoid ambiguity and safety

requirement decomposition pattern which is model-based structural guideline for expression of safety requirements.

Above methods cannot be used directly to solve problems what we focus. The limited ability regarding accuracy and evaluation scope is the main constraint of automated approaches. In case of review support method, reviews of research artifacts are not considered. Also, reviews which performed by untrained reviewer are not considered.

3 Content-Based Conformance Assurance

We take following approaches for assurance of conformance between design guideline and software research documentations. Since most of software standards that reflect best practices of software development requires requirements traceability, our method takes model-driven traceability approach. However, we extend traditional traceability to semantic-aware traceability for trace between research documents and development documents. Also, we develop a design guideline extracted from software standards for provide knowledge to our method. Further, we transformed the guideline into semantic rules and checking conformance of guideline in automated manner for reducing efforts of assessment. Finally, we present a series of explanatory guidelines which explains rationale which shows the source of the metrics to users to give better understanding of the best practices.

3.1 Relevance Link Information Model (RLIM)

Traceability is an ability to establish links between source artifacts and target [1]. This attribute is essential for every software standard. The existing traceability mainly focus on artifacts which include requirements. Also, a trace link means transitive relation while research documents have non-requirement contents and non-transitive relations [4]. We extend trace link to Relevance Link (RL) in [4] these limitations. A RL composed with 2 major components. Corresponding Items is two configuration items having relevance and Relevance Rule is a rule which corresponding items should follow. We defined 7 relevance rules in [4]. Therefore, A Relevance Link becomes a mapping between the corresponding items and relevance rule.

3.2 Expert Assessment Goal Model

We propose an Expert Assessment Goal Model which aims to representing knowledge for best practices of software development and direct measurement of conformance. We created goal models from three domains, each goal model guidelines based on international software standards, which are IEC 12207, ISO 26262 and IEC 62304. This guideline structure is combined structure of Goal, Question, Metric model [6] and Goal Structuring Notation [10] for satisfying both goals. Figure 1 shows the detailed goal model structure.

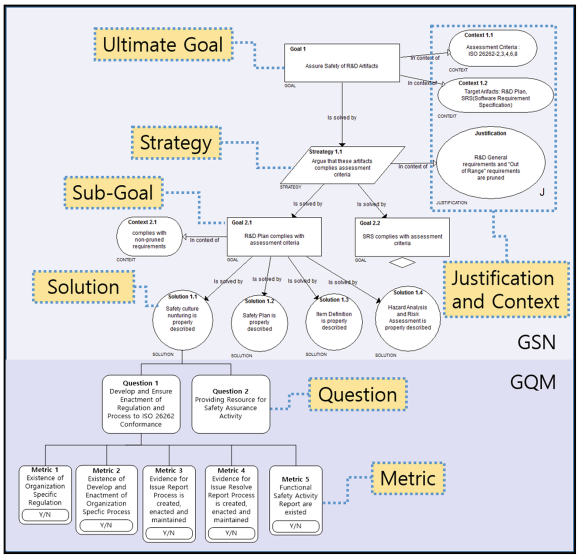


Fig. 1. This shows the goal model structure of Expert Assessment Goal Model. *Ultimate Goal* is a quality that documents should achieve and divided by *Sub-Goal*s. Each *Sub-Goal* is divided by *Solutions*. Below a *Solution,* there is a *Question* and a *Metric* for measurement of solution. There also are *Justification* and *Context* to showing the adequacy of a goal model to standards. Additionally, we annotated the source of each *Metric*.

3.3 Conformance Assurance

We utilize a combined method of above methods to check conformance. First, we bridge the gap between RLIM and Expert Assessment Goal Model by translating the goal model into rules that contain relevance rules. The translations can be done differently by the scope of measurement. Next, RLIM-based expert system [5] checks conformance using rules and document RLIM. This system utilizes a rule engine which can evaluate the conformity of rule therefore we can check the conformance of design guidelines as rules with this system in automated manner. However, in case of metrics which needs semantics analysis, contents need be evaluated with other review systems that can evaluate semantics.

4 Application

We conducted a preliminary experiment to verify effectiveness of our method from our previous project. First, we transformed the metrics on the guidelines into rules. Each metrics can be defined as a query for certain relevance links or contents of documents. Figure 2 shows some examples of transformed guidelines.

Next, documents of interest are should transformed into RLIM. In this paper, we let developers to manually build document RLIMs for the best accuracy of the assessment. Then we conducted the assessment using RLIM expert system. The results of assessment displayed as shown in Table 1. Also, the explanatory guidelines given as Fig. 3.

Guideline	Conformance Criteria	Trace Query for assessment (1st order logic)
Internal Consistency of SW Reqs.	If there is a trace link between two items, two items must have no conflict	∃t1∃t2∃rr (Req.(t1)∧Req.(t2)∧Rel.Rule(rr) ∧ RLIM(t1,SRS,t2,SRS,rr) ∧ Conflict(t1, t2))
Integrity between Research Goal and SW Reqs.	Research Goal and at least 'Threshold'% of SW requirements should have semantic sufficient (SE-SC) condition.	∃t1∃t2 (Research Goal(t1) ∧ Req.(t2) ∧ RLIM(t1,SRS,t2,SRS,SE-EC))
Conformity between Research Goal and Plan of Action	Make sure that corresponding items have at least semantic sufficient (SE-SC) condition	∃t1∃t2 (Research Goal(t1) ∧ Plan of Action(t2) ∧ RLIM(t1, Research Plan, t2, Research Plan, SE-SC))

Fig. 2. The example of transformed guidelines into trace query. RLIM(*t1, ci1, t2, ci2, rr*) means a relevance link which has corresponding item *t1* from *ci1* and *t2* from *ci2* with relevance rule *rr*. *Ele*(x) means the element type of x is '*Ele*'.

Table 1. The result of our preliminary application. Users receive an evaluation report in this form. We found that documentation problems of risk management process on Software R&D Plan. Also, we found that our Software Requirement Specification meet very little criteria.

Target document	Evaluation criteria	Conformance score
Software R&D Plan	Research goal, motivation, trends	Acceptable (67 %)
	Organization description	Good (88 %)
	Research strategy	Good (83 %)
	Risk management process	Poor (0 %)
Software Requirement Specification	Software requirement description	Poor (22 %)
	Requirement analysis	Marginal (33 %)
	Software test planning	Poor (0 %)

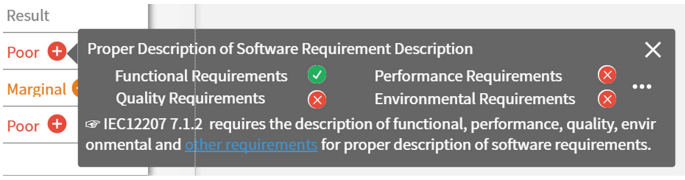


Fig. 3. This shows an explanatory guidelines of the evaluation results. The detail of assessment is provided to user for a basis of useful feedback.

5 Conclusion

In this paper, we propose a content-based method of conformance assurance between software research documents and design guideline. We first extract design guideline from software standards. Then, we transformed guidelines into rules for automated evaluation. We also used an expert system which can evaluate the conformance of rules

transformed from guidelines. With our method, we expect the improvement of document quality of research groups who need to develop software. In future works, we will improve the guideline for more helpful and build practical applications for documentation model and evaluation.

Acknowledgement. This research was supported by Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF- 2014M3C4A7030504).

References

1. Aizenbud-Reshef, N., Nolan, B.T., Rubin, J., Shaham-Gafni, Y.: Model traceability. IBM Syst. J. **45**, 515–526 (2006)
2. Allman, E.: Managing technical debt. Commun. ACM **5**, 50–55 (2012)
3. Antonino, P.O., Trapp, M., Barbosa, P., Sousa, L.: The parameterized safety requirements templates. In: Proceedings of the 8th IEEE/ACM International Symposium on Software and Systems Traceability, pp. 29–35 (2015)
4. Baek, D., Lee, B., Lee, J.W.: Content-based configuration management system for software research and development document artifacts. KSII Trans. Internet Inform. Syst. **10**, 1404–1415 (2016)
5. Baek, D., Shin, J., Lee, B., Lee, J.: Toward development of a traceability model measuring compliance with guidelines. In: 11th KSII Asia Pacific International Conference on Information Science and Technology, pp. 37–38 (2016)
6. Basili, V.R.: Software modeling and measurement: the Goal/Question/Metric paradigm (1992)
7. Dautovic, A., Plosch, R., Saft, M.: Automatic checking of quality best practices in software development documents. In: Proceedings of the International Conference on Quality Software, pp. 208–217 (2011)
8. Goble, C.: Better software, better research. IEEE Internet Comput. **18**, 4–8 (2014)
9. Jain, P., Verma, K., Kass, A., Vasquez, R.G.: Automated review of natural language requirements documents: generating useful warnings with user-extensible glossaries driving a simple state machine. In: Proceedings of the 2nd Annual Conference India Software Engineering, pp. 37–45 (2009)
10. Kelly, T., Weaver, R.: The goal structuring Notation–A safety argument notation (2004)
11. Shen, W., Lin, C.L., Marcus, A.: Using traceability links to identifying potentially erroneous artifacts during regulatory reviews. In: Proceedings of the 7th International Workshop on Traceability in Emerging Forms of Software Engineering, pp. 19–22 (2013)
12. Thitisathienkul, P., Prompoon, N.: Quality assessment method for software requirements specifications based on document characteristics and its structure. In: Proceedings of the 2nd International Conference Trustworthy Systems and Their Applications, pp. 51–60 (2015)
13. Wilson, W.M., Rosenberg, L.H., Hyatt, L.E.: Automated analysis of requirement specifications. In: Proceedings of the International Conference on Software Engineering, pp. 161–171 (1997)

Development of the Vision System and Inspection Algorithms for Surface Defect on the Injection Molding Case

Ji Yeon Lee, Wonwoo Bong, Sangjoon Lee, Chang Ho Han(✉), and Kuk Won Ko

School of Mechanical and ICT Convergence Engineering, Sunmoon University, Galsan-ri, Tangjoong-myoon, Asan, Chungnam, Korea
jylee2930@gmail.com, dmenl0@naver.com, {mcp94lee, liberman, kuks2309}@sunmoon.ac.kr

Abstract. The surface defects of plastic after the injection molding process have been inspected human inspector. In this paper, we propose the automated surface inspection system to detect 3D defects on injection molding objects. This vision system is equipped with 4-cameras and 8-illuminations and is designed to find optimal configuration between camera, illumination and object to detect the defects on an object. Also we develop algorithms to detect defects in captured images. The developed surface inspection system reduce the labor costs and time, whereas increase the detection accuracy of the surface defects such as scratch. The proposed algorithm uses the difference of the brightness of the obtained image by the camera and illumination system for acquiring data of the surface of the object. We focus on the surface defect detection of black molding object. Scratch detection is performed by the developed vision inspection system.

Keywords: Non-flat surface · Defect detection · Scratch · Multi-camera · Multi-illumination

1 Introduction

In general, the injection molding process is one of the plastic molding method that cooling solidified while maintaining in a mold at a high pressure after the heat melting plastic. In the injection molding process, surface defects such as scratches, nicks, mark broken, and pollution. Because these defects detection usually proceeds by hand, it is difficult to obtain a uniform quality determination criteria for the bad checks between the operators. Because the operators are investigation the product to rotate several angles, it is very time-consuming and the inspection is inconvenient task. For reducing the inspection time and for increasing the inspection accuracy, we propose the automatic inspection vision testing system. Inspection vision testing system is that the rotation stage is mounted on a frame of the dome to rotate for the camera and the LED illumination [1, 2] and the object. The selecting of different angles of illumination and camera for achieving the same effect as the human inspector is very important.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_150

In this study, we selected camera angles and lighting to optimize defect detection and developed detection algorithm for the defects [3–6].

2 Vision Inspection Systems

Looking at the manual's inspection method, in order to look good the defects will be inspected while turning the object from the angle of a variety of lighting and operator visibility. This is because easily visible defects or invisible defects are depending on the angle of illumination, eye and the object. Therefore, in order to detect the various defects, the operators must be able to find the relative position between the inspection object, the optimal camera and the illumination easily. Development vision inspection system includes cameras and illuminations of various angles, to be able to capture the images of the object, as in Fig. 1. It established a tunable illumination to 0–90° domed frames to maintain the camera and the illumination and the object a certain distance, and so also fixed at right angles. The camera improves the inspection speed, to inspect from various angles, designed to set up a four.

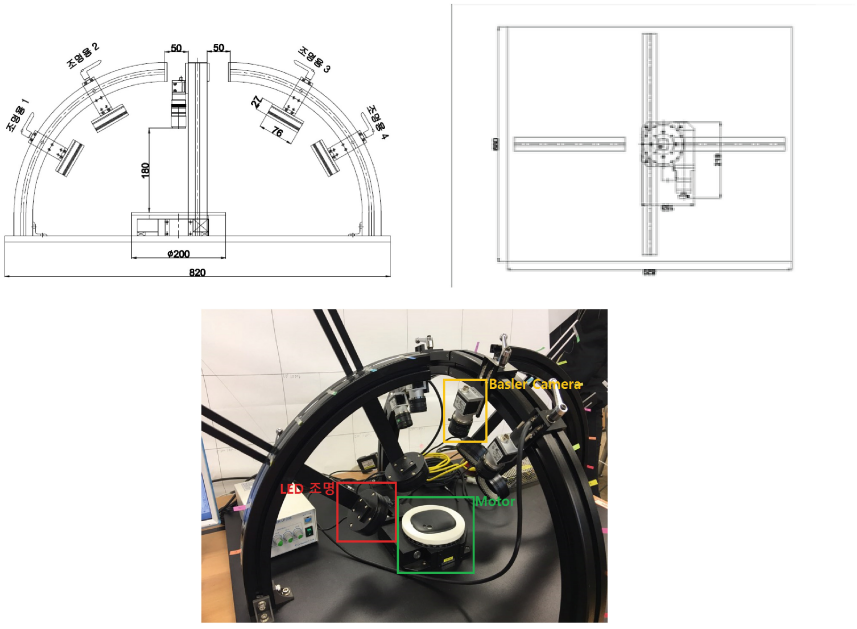


Fig. 1. Vision inspection system

2.1 Camera and Rotation Stage

We set up a 4 Basler camera to see the object from different angles. The rotation stage that is designed to get all of the images of object by rotating 360°. The rotation stage use the Fastech Motor and Ezi-Servo and EzM-42S-A-D driver, and the diameter of the

stage is 42 mm, the length of the motor is single, the encoder resolution is 10,000/rotation. Specification of Basler camera and Fastech Motor for getting a picture of the product are shown in Table 1. Figure 2 shows the Basler camera and the rotation stage.

Table 1. Specification of camera and motors

Basler camera		Fastech motor	
Product name	acA1300-30μm (mono)	Product name	Ezi-Servo Plus_R
Resolution	1.3 M	Resolution (P/R)	10,000/rotation
Pixel size	3μm∼4μm	Input voltage	24VDC ± 10 %
Frame rate	21∼50	Rotation speed	0∼3,000 rpm
Interface	USB3.0		
Sensor size	4.86 mm × 3.62 mm		

Fig. 2. Basler camera and rotation stage

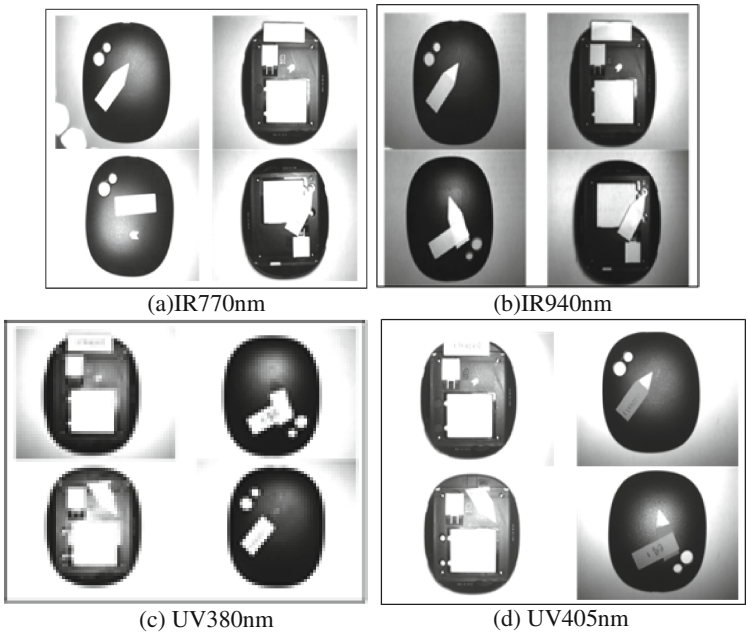


Fig. 3. Defect detection by LED wavelength

2.2 LED Illumination

We set up 8-circular POWER LP 205 LED illuminations to see the object from different angles to have redundancy of light configurations. The wavelength of illumination is important to emphasize the gray level difference of the defects on normal surface. We experiment on the possibility of defect detection according to the wavelength, the results are shown in Fig. 3. Experimental result shows that as the wavelength of illumination is closer Ultraviolet wavelength, the defect detection is clearer.

3 Surface Defect Detection

3.1 Optimize the Angle of the Illumination and Camera

In order to find the optimal angles of camera, illumination and motor, after the angle of the illumination is fixed, the angle of the camera changes 10° steps to 90° from 0°, while the angle of the motor changes 20° steps to 360° from 20°. Each different configuration, when we detect the scratches. We marked red and yellow, in Fig. 4. The optimal angles of camera are from 70° to 90°. Illuminations are 70°, 90°, from 140° to 150°. Motors are from 40° to 120°, from 180° to 240°. The Fig. 4 is the case of the 90° angle of the illumination. While the angle of the illumination changed 10° from 10° to 80°, we repeat the experiment of the experiments to find the optimal optical configuration of our vision system (Fig. 5).

Angle of Illumination: 90

Angle of Cam	10°	20°	30°	40°	50°	60°	70°	80°	90°
20	0	0	0	1	1	2	2	2	0
40	0	0	0	0	0	2	2	2	1
60	0	0	0	0	1	2	2	1	1
80	0	1	0	2	2	2	2	1	1
100	0	0	0	2	2	2	2	1	2
120	0	0	0	2	2	2	2	1	2
140	0	0	0	1	1	2	2	2	2
160	0	1	1	2	2	2	2	2	2
180	0	1	1	1	1	1	2	2	2
200	0	0	0	1	1	1	2	2	2
220	0	0	0	1	1	1	1	2	2
240	0	0	0	1	1	1	1	2	1
260	0	0	0	0	1	1	1	1	2
280	0	0	0	0	1	1	1	1	2
300	0	0	0	0	1	1	0	2	2
320	0	0	0	0	0	1	1	2	2
340	0	0	0	0	0	1	2	2	2
360	0	0	0	0	0	1	2	2	2

Fig. 4. Result of defect detection by various angles

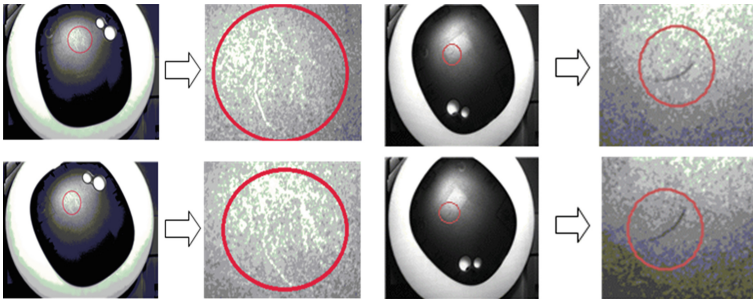


Fig. 5. Scratch appears by various angles

3.2 Algorithm for Surface Detect

Conventional detecting scratch of surface algorithm is to use binary threshold technique [7]. This method is useful when the surface of object is flat and its gray value is regular. However, our target has rounded surface and non-flat gray level value of surface. The other method is edge detection based on Sobel or Laplacian operators. Scratch defects may be shown as edge in image. This edge based method is also not effective because surface are irregular shape and surface have irregular gray value at round part of surface. These two methods are difficult to apply to objects having a non-uniform surface and it make irregular brightness at a various illumination configurations.

In our proposed algorithm, we have the reference image that is stored and smoothed. Smoothing can removes noise caused by the brightness change. This good original image is used as reference image. The Scratch image is compared with smoothed reference image and scratch defects are detected by simple thresholding algorithm as shown in Fig. 6. To avoid detection errors, the offset value should carefully selected (Fig. 7).

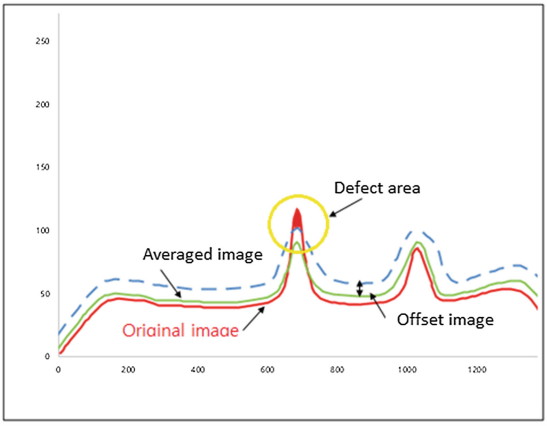


Fig. 6. Scratch detection algorithm

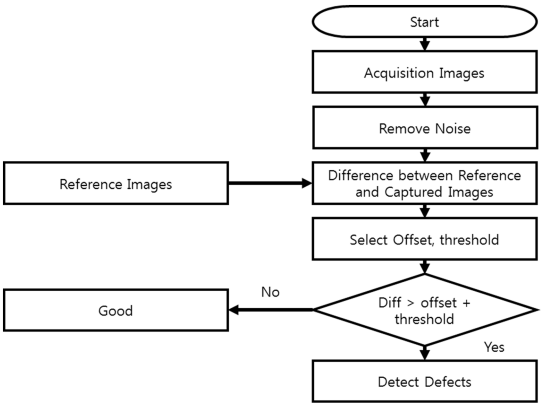


Fig. 7. Flowchart of surface scratch detection

4 Experiment and Result

Figure 9 shows the developed our program UI to display all images obtained from 4 cameras with MFC. The 4 images in the first row are the reference images obtained with golden samples, the 4 images in second row are the obtained images at various configurations, the 4 images in third row are the result of the images after image processing. Proposed algorithm are easily explained by images and graphs of gray level on the detection in Fig. 10. Green line means filtered gray level of golden objects and red line is filtered gray level of captured image. Blue mark is detected defect on the scratch. Table 2 shows the inspection results with our proposed system comparing to human inspectors. Experiment result shows 98.7 % with better accuracy compared with human inspectors (Fig. 8).

Table 2. Specification of camera and motor

Inspection method	Number of trials	Number of successes	Number of real defect	Ratio (%)
Human inspector 1	1000	971	29	97.1
Human inspector 2	1000	982	18	98.2
Human inspector 3	1000	978	22	97.8
Proposed method	1000	987	13	98.7

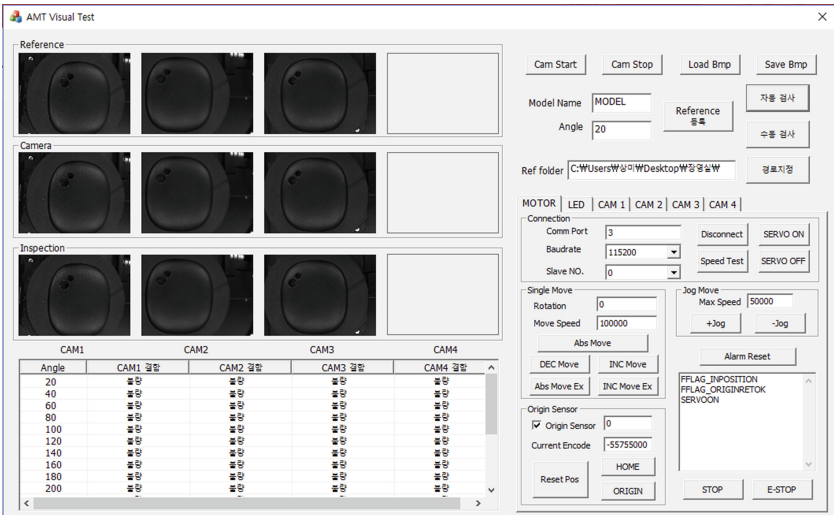


Fig. 8. Graphic of user interface

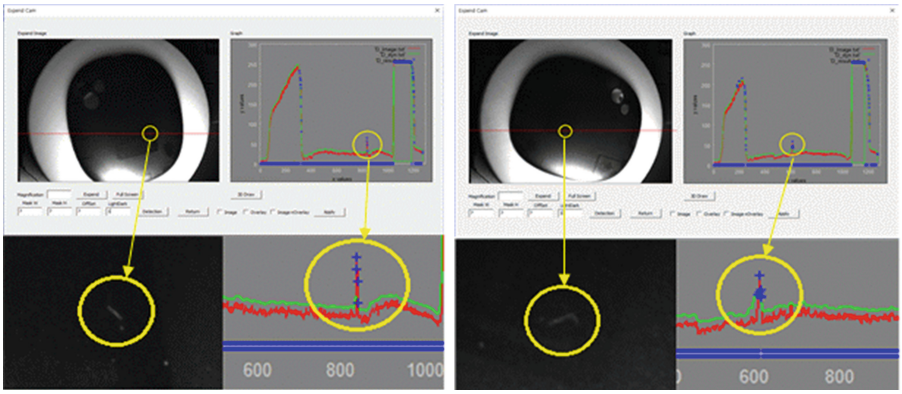


Fig. 9. Graphic of user interface

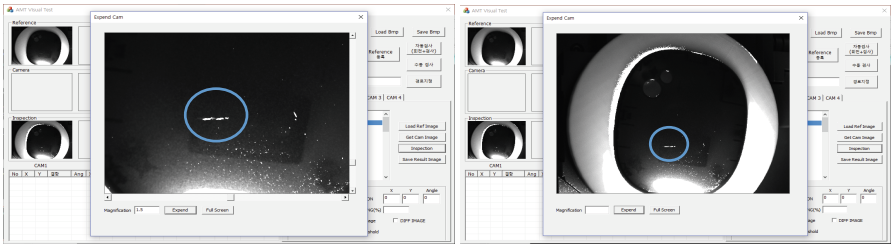


Fig. 10. The result of scratch detection

5 Conclusion

A surface inspection system for defects on injection molding case is presented based on multi-camera, multi-illumination to capture images of object around and the proposed subtraction image processing algorithm between reference and captured images. In this study, experiments are for the black molding object. It is demonstrated that the detecting works for tiny and invisible defects on surface. We show the result in the ability of detection in non-flat objects with the proposed algorithm. In the future, the applicability of detection of complex shaped object with various color will be considered. In the future, our study will concentrate on developing image acquisition system to obtain better quality finger-vein images and PPG signals and new fusion algorithms for high success rate and fast identification.

Acknowledgement. This research was financially supported by the Human Resource Training Project for Regional Innovation and Creativity (NRF-2014H1C1A1066998).

References

1. Stark, W.S., Tan, K.E.W.P.: Ultraviolet light: photosensitivity and other effects on the visual system. Photochem. Photobiol. **36**(3), 371–380 (1982)
2. Ryer, A.: Light Measurement Handbook. International Light, Newburyport (1997). http://www.intl-light.com/handbook/index.html. Accessed 19 Sep 2000
3. Bourgeat, P., Meriaudeau, F., Gorria, P.: Defect detection and classification on metallic parts. In: Proceedings of the SPIE Machine Vision Industrial Inspection X, San-Jose, USA, vol. 4664, pp. 182–189, January 2002
4. Shih-Chieh, L., Chih-Hsien, C., Chia-Hsin, S.: A development of visual inspection system for surface mounted devices on printed circuit board. In: Proceedings of the 33rd Annual Conference of the IEEE Industrial Electronics Society Taipei, Taiwan, pp. 2440–2445 (2007)
5. Kang, S.M., Park, S.H., Huh, K.M.: Development of remocon appearance inspection system using automated vision. In: IEEE Summer Conference, vol. 6, pp. 695–696 (2006)
6. Kim, J.H., Ko, K.W.: Research on defect inspection of LCD panel. J. Korea Soc. Precis. Eng. **24**(4), 7–11 (2007)
7. Kang, D.J., Ha, J.E.: Digital Image Processing. Infinity Books, Korea (2010)

Implementation of the Smart System for Monitoring the PCG

Sunho Kim, Kangwoo Lee, and Yonghee Lee(✉)

Department of Computer Engineering, Halla University,
Wonju-shi, Kangwon-do, Korea
yhlee@halla.ac.kr

Abstract. In this paper, we implement the smart system for monitoring the phonocardiogram of a patient. The proposed system consists of a module to measure the phonocardiogram with Bluetooth communication and an encoder module for storing the measured phonocardiogram to a standard protocol. The system enables real time monitoring while following the international standards to store the health signals. Especially, we implement the system to monitor the PCG, and we show that the smart system can be utilized to monitor phonocardiogram of patients in a mobile environment.

Keywords: Mobile · PCG · Phonocardiogram · Encoder · Smart · Healthcare

1 Introduction

The cardiac sounds are ones to be recorded by waves after switching sound energy to electric energy and called PCG(phonocardiogram). The examination of cardiac sounds is generally implemented by auscultation, but the cardiac sounds can be analyzed objectively and accurately comparing to auscultation and can record the sounds related to the heart. And it has a merit of improving the auscultation ability by comparing phonocardiogram and auscultation opinion. But the PCG can never be a substitution of auscultation. It proposes more minute opinion by being looked at and listened to together the PCG and auscultation. In any case, examination with either only auscultation or phonocardiogram is incomplete. The cardiac sounds can be normally judged by considering the relationship with ECG(electrocardiogram). Encoding methods for utilizing the waves like this PCG with variety in the medical environment of health care, remote medical treatment and health monitoring system have been developed, saved and transmitted autonomously by each company. That is, the specifications for storage and transmission follow manufacturers' specifications, which makes compatibility between systems of different equipment to be difficult. And most system require PC environment, which limits mobility and portability.

Lately MFER(medical waveform encoding rules) method was enacted and is used as International Standard for saving effectively waveforms information used at clinical diagnosis. The MFER can be easily realized by a user, saved as a file unit, and used simply and effectively. The merit of this encoding method is to easily realize due to its simple rules. But it has a structure inappropriate to monitor real-time because of file

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_151

unit encoding. That is, the storage of MFER as a file unit prevents a user from finding the contents of PCG before completion of measurement, and from transmitting before saving. And also the method limits portability and mobility because of its operation in the PC environment. This study proposes an encoding structure which enables to save and transmit by a certain data size in order to solve these difficulties. And this proposes a PCG encoding system which can be effectively utilized under the smart environment so as to make good use for remote clinical treatment with high mobility and portability.

2 Methods

2.1 Measurement of Cardiac Sounds

PCG is used for making a diagnosis of patient through auscultation other than ECG. A cardiac sound measuring apparatus is used for comparing and quantitatively evaluating by acoustically recording the movement of heart. Even though the examination of cardiac sounds is executed with auscultation, the phonocardiogram can be objectively analyzed with detail comparing to auscultation and has an advantage that can record with accuracy the low frequency sound close to the low limit of normal auscultation (20–20,000 Hz/s). And also the comparisons of phonocardiogram and auscultation opinion enable to upgrade the auscultation ability. The normal phonocardiogram has four sounds of I, II, III and IV shown at Fig. 1. I and II sounds are called as cardiac sounds for contraction period, and III and IV are called as cardiac sounds for diastole. I sound is one occurred at the early of contraction period, II sound is one occurred by closure of semilunar valve at the end of contraction period. III sound is occurred by vibration of ventricular screen when the blood flows into the ventricles from atrium at the abundance period. IV sound (or atrium sounds) is low and dull one occurred after 0.12 s from the beginning of P wave and does not sound to normal people.

The measurement of cardiac sounds is executed around the measuring zone on the chest like Fig. 2. The PCG signal input in the pre-treatment process shall be filtered in order to remove artefact and noises. The band of cardiac sounds is 10 Hz–1 kHz, which artefact and noises are removed through the pre-treatment. The measured PCG

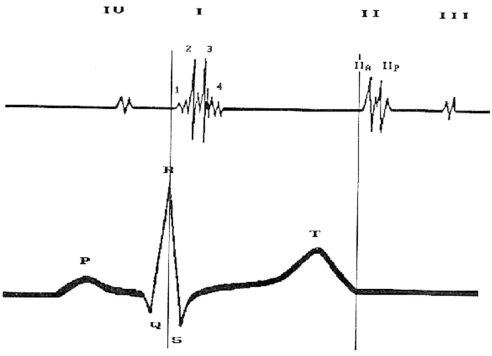


Fig. 1. Phonocardiogram

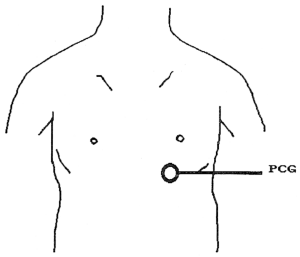


Fig. 2. Measurement of PCG

signals are switched to digital signals at the A/D switching zone through amplification zone. The switched signals go through filters according to features of signals. And they are converted as a file type by encoding on standard according to a encoding rule in MFER encoding zone.

2.2 Standardized Encoding Part

The International Standard for encoding of bio-signal was enacted at ISO/IEEE11073. Lately MFER standard was enacted in order to encode the waveform information used in the medical section such as PCG. It is used maintaining the compatibility of bio information in the diverse sections. The basic structure of MFER is composed of a header and waveform data, which have tag, length and value. Tag (T) represents property of data value, Length (L) does the length of data value, and value (V) shows the contents represented by Tag (T). Figure 3 shows encoding rule of TLV type.

Fig. 3. TLV format for encoding medical signals

When MFER sends the medical waveform data, it must send as a file type after completion of measurement. Therefore it cannot get the waveform information during the measurement. It is difficult to check before the completion of measurement when the data need to be transmitted or monitored during the measurement. This study proposes the encoding like Fig. 5 in order to solve these problems.

2.3 Smart Encoding System

The proposal system in Fig. 4 basically follows MFER, International Standard, but can save and transmit files during the measurement. Figure 4 shows the existing MFER file structure. It has a structure to make one file by combining a header and a waveform data. Figure 4 shows a structure of dividing the medical waveform data into certain

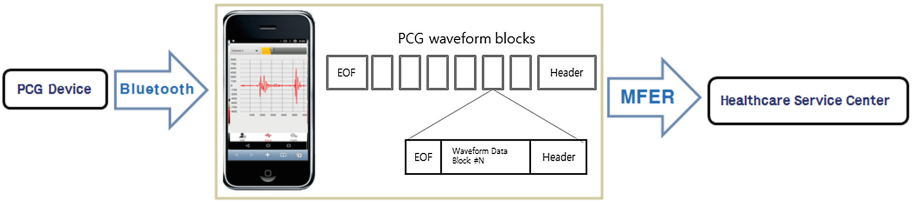


Fig. 4. Proposed encoding apparatus

data sizes as separate files and saving the files. When the size of data is small, it can be saved while monitoring because the difference between measuring time and saved time is small. Several data files with certain length come out. Finally a file can be made by combining the header and data information.

3 Test and Results

As Fig. 5, when clicking a file with.mwf extension, the Info screen shows Header information of file, and the Wave screen displays the actual waveform. It is composed of Info, Wave and From menus. When we read a file, it processes Info menu and Wave menu in order. And when we use Bluetooth communication, the Head information in Info Menu is saved first, and then saves Waveform data transmitted in real time by

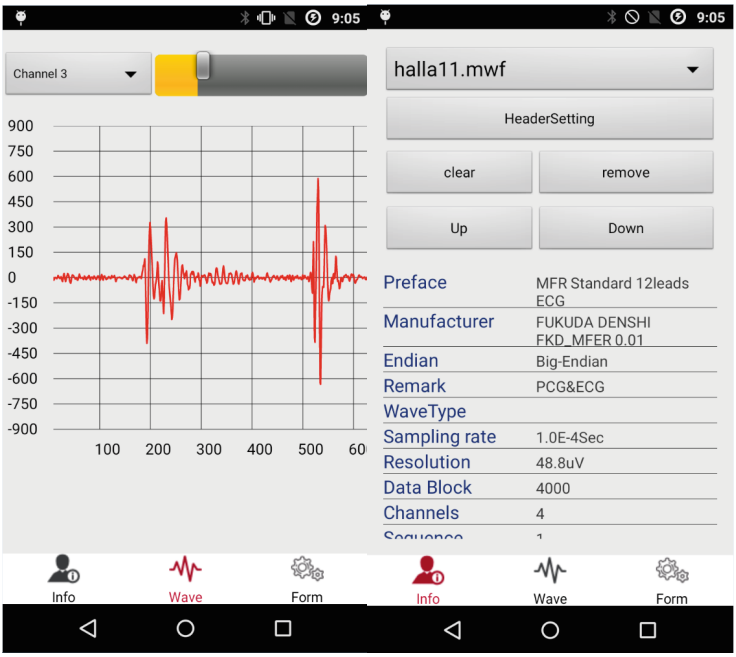


Fig. 5. Smart encoding system

combining them to headers. This can prevent loss of data from completion of Bluetooth communication or communication error. The activity of reading file is executed Form menu, Info menu and Wave menu in order. When clicking an item on the file list in Form menu, it is transmitted to Info List Method. Info List processes Header information and then Wave Viewer Method in Wave menu processes Wave Form Information. The rule processing values follows MFER rules. Waveform process is executed in WaveActivity class. When InfoList Method in InfoActivity processes Header, it transmits block size, number of repeated file, number of channel, endian, unit size of waveform, file name, properties of waveform to Wave Viewer Method and outputs waveform information. The whole length of the waveform is calculated as block size * number of channel * number of repeating * data size.

4 Conclusions

This study suggests an encoding system under the smart environment in order to save and transmit the cardiac sounds. The smart PCG monitoring system has an MFER structure, which enables to save and monitor by maintaining compatibility with the existing system. The system has mobility and portability by realizing as an application of smart phone. It can save in real time without a big increase of data, which would be effective in the remote medical treatment.

References

1. Hirai, M., Kawamoto, K.: MFER -a Japanese approach for medical wave form encoding rules for viewer design. In: OpenECG Workshop 2004, Berlin, Germany (2004)
2. Kimura, E.: Development MFER (Medical waveform Format Encoding Rules) Parser. In: Amia Annual Symposium Proceedings-CD-Rom Edition (2006)
3. Lee, Y., Kim, S.S., Choi, C.H., Park, J.H.: Encoder design for healthcare signals. Pers. Ubiquit. Comput. **17**, 1373–1381 (2012). Springer
4. James, A.P.: Heart rate monitoring using human speech spectral features. Human-centric Comput. Inf. Sci. **5**, 33 (2015)

The Effect of Introducing Small Cells in Wireless Networks

Soohyun Cho(✉)

Hongik University, 94 Wausan-ro, Mapo-gu, 04066, Seoul, Korea
cho.soohyun@hongik.ac.kr

Abstract. The use of small cells is one of the key elements for future wireless networks, which face ever-increasing demand for higher capacity. In this paper, we investigate the effect of introducing small cells in IEEE 802.11 wireless networks where access points of low transmit powers are mixed with access points of normal transmit powers. We use Poisson point processes to randomly locate normal access points, low-powered access points for small cells, and wireless users in a given area. We evaluate the effect of introducing small cells in the point of fairness among the wireless network users through simulations.

Keywords: Small cell · Voronoi diagram · Poisson point process

1 Introduction

To cope with the ever-increasing demand of wireless network capacity, heterogeneous networks (HetNets) are considered as one of the key solutions for future mobile networks [1]. HetNet wireless networks are composed of heterogeneous cells (i.e., mixed with macro cells and small cells). Macro cells usually contribute to building a wireless coverage in which a minimum throughput for mobile communications is guaranteed (usually for voice communications). Small cells (also known as femtocells or picocells) are expected to contribute to building hotspot zones which can offload the load of busy macro cells covering large areas.

HetNets are, however, mainly considered as key technologies for future mobile wireless networks such as the 5 G network [2]. Other wireless networks such as IEEE 802.11 wireless networks could also introduce small cells to improve the capacity and performance. Small cells, however, may result in degradation of overall network performance when they are introduced in IEEE 802.11 wireless networks without careful study of the impact they might bring in. This is because IEEE 802.11 wireless networks have quite different operation mechanisms from the operator-managed wireless mobile networks such as LTE.

Basically IEEE 802.11 wireless networks are based on the cooperation among participating mobile nodes. The request to send/clear to send (RTS/CTS) mechanism and the random backoff in the carrier sensing multiple access with collision avoidance (CSMA/CA) scheme are examples of the cooperation. But, one of the most important assumptions behind is that participating entities use the same transmit power not to unfairly utilize wireless resources. Introducing small cells by deploying access points

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_152

(APs) of less transmit power than that of normal APs may result in unfairness to wireless network users who associated with small-cell access points.

In this paper we focus on the effect of introducing small cells in IEEE 802.11 wireless networks at infrastructure mode. We use the ns-2 [3] network simulator of version 2.35 to evaluate the performance and the fairness among users of heterogeneous IEEE 802.11 wireless networks.

2 User Associations and Voronoi Diagram

We first investigate the associations between wireless users (or mobile nodes) and access points of IEEE 802.11 wireless networks and compare them to Voronoi diagrams [4]. Throughout this paper, we use independent Poisson point processes of different densities to randomly select the positions of mobile nodes (MN), normal-cell access points (AP), and access points for small cells (SAP) in a given area. With homogeneous networks (i.e., all access points use the same transmit power), the association between mobile nodes and access points can be analyzed from the positions of the mobile nodes and the access points [5].

In Fig. 1, we show simulation results of the associations between 10 mobile nodes and 10 access points using the lines between the access points and the mobile nodes associated with them. In the simulations, all access points and mobile nodes use 100 mW for transmission. Figure 1 also shows the tessellation of ordinary Voronoi diagram constructed using the locations of the access points in the plane. The figure shows that all the associations between the mobile nodes and the access points accurately coincide with the Voronoi areas of access points.

The accuracy comes from the policy with which the mobile nodes make associations with the access points. In our simulations, each mobile node calculates received signal powers (P_r) of beacon signals from each access point using Eq. (1) from Friis [6] and makes association with the access point of the strongest received signal power.

$$P_r = P_t G_t G_r \lambda^2 / ((4\pi d)^2 L). \tag{1}$$

Here, P_t is the transmit power of access points, and G_t and G_r is the antenna gain of the transmitter and receiver, respectively. λ is wavelength of the signal, d is the distance from a mobile node and an access point, and L is a loss parameter.

For the heterogeneous network cases (i.e., wireless networks composed of access points of different transmit powers), we use a similar approach to investigate the associations between mobile nodes and access points of different transmit powers: we first simulate the associations and compare the results to Voronoi diagram. Figure 2 shows the association results when we change the transmit powers of 5 access points in Fig. 1 from 100 mW to 10 mW. The figure also shows the multiplicatively weighted Voronoi diagram constructed using the locations of the access points and their transmit powers.

In multiplicatively weighted Voronoi diagrams, tessellations are decided not only by the geographical distribution of the center of each cell but also by the weight of each cell with the following multiplicatively weighted metric:

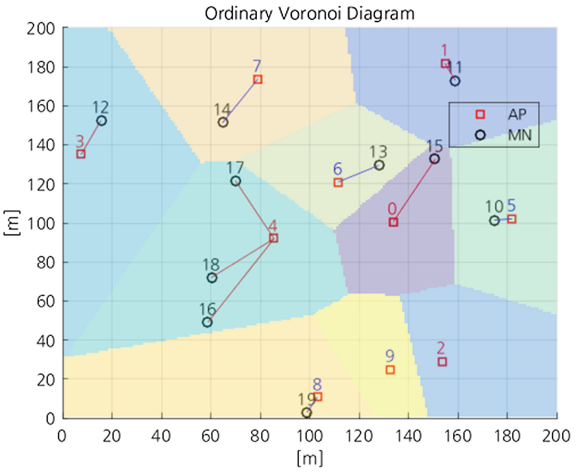


Fig. 1. Ordinary Voronoi diagram and associations of 10 mobile nodes (MNs) and 10 access points (APs) in an area of 200 m × 200 m.

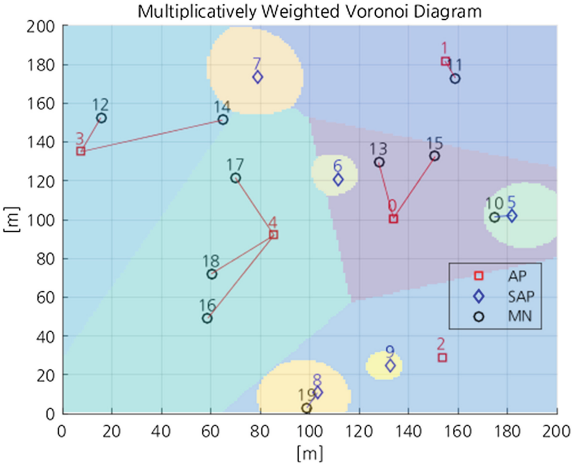


Fig. 2. Multiplicatively weighted Voronoi diagram and the associations of 10 mobile nodes with 5 normal APs and 5 small-cell APs (SAP) in an area of 200 m × 200 m.

$$\| x - x_i \| / w_i < \| x - x_j \| / w_j. \tag{2}$$

Here, x is a point in a plane. w_i and w_j are the weight of cell i and j, respectively. x_i and x_j are the position of the center of cell i and j in the plane, respectively.

However, if we use the transmit power of each access point (p_i) as its weight (w_i) in Eq. (2), the resulting tessellation is not accurate compared to the association results from simulations. We need to use the square root of the transmit power of each cell as

its weight since the distance d in Eq. (1) is squared. This can be accomplished by expressing the metric in Eq. (2) as below:

$$\| x - x_i \| / \sqrt{p_i} < \| x - x_j \| / \sqrt{P_j}. \tag{3}$$

The tessellations in Fig. 2 show that we can accurately describe the associations between mobile nodes and access points of different transmit powers with the metric in Eq. (3) when Eq. (1) is used as the radio propagation model.

3 Simulation Results

To evaluate the performance of the heterogeneous IEEE 802 wireless networks, we use and develop the following features in ns-2 version 2.35:

(1) To investigate association relationships between mobile nodes and access points, we first use the infrastructure mode of IEEE802.11. Here, `RXThresh_` and `CSThresh_` parameters are set to very small values to allow all participants to detect each other.
(2) After gathering the association relationships, for actual simulations, we use the 802.11Ext[1] module from [7] and parameter recommendations for IEEE802.11a [8] to investigate performances with different modulation schemes. The module supports four modulation schemes (i.e., BPSK, QPSK, 16-QAM, and 64-QAM) to designate the minimum signal-to-interference-plus-noise ratio (SINR) requirements for proper packet decoding.
(3) We adopt No Ad-hoc Routing Agent (NOAH) feature from [9] to designate fixed and direct routing paths between access points and their associated mobile nodes using the association relationships.
(4) We modify the packet propagation scheme in ns-2 (i.e., `channel.cc`) related to the 802.11Ext module for access points of different transmit powers to have different propagation ranges: only mobile nodes in their propagation ranges receive packets.

In our simulations, access points are continuously backlogged by constant-bit-rate (CBR) traffic of 10 Mbps to each mobile node associated with them from one second after starting simulations. The packet size of CBR traffic is 1,000 bytes and the queue size for the application traffic is set to 50 packets. The RTS/CTS scheme is disabled and TwoRayGround model [10] is used for radio propagation. All the other parameters for simulations are left to the default values of the ns-2 distribution.

Figure 3 shows the changes of throughputs of 10 CBR flows from access points to their associated mobile nodes during the first three seconds of simulations. All flows use the QAM64 modulation scheme. Figure 3(a) is the simulation results of the homogeneous case depicted in Fig. 1. Figure 3(b) is the simulation results of the

[1] Since the 802.11Ext module does not support the infrastructure mode, we use NOAH to explicitly disable ad-hoc routings.

heterogeneous case depicted in Fig. 2. Since the size of the area (i.e., 200 m × 200 m) used for simulations is small compared to the transmit powers of access points, all access points detect traffic from each other including small-cell access points and control their sending rates according to the CSMA/CA scheme.

As a result, as Fig. 3(b) shows, small-cell users (i.e., MN 10 and MN 19) eventually achieve more per user throughput than normal-cell users because the number of users associated with small cells are smaller than that of normal cells as the Voronoi diagram in Fig. 2 suggest.

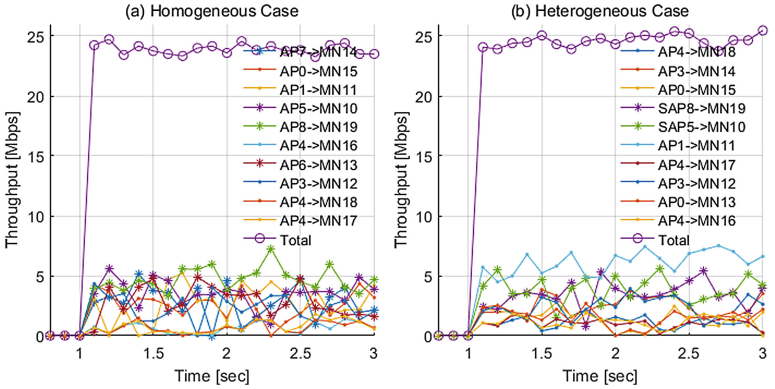


Fig. 3. The changes of achieved throughputs by 10 MNs: (a) of Fig. 1, (b) of Fig. 2.

However, when we increase the size of the area, the per user throughput of wireless users associated with small-cell access points and that of wireless users associated with normal-cell access points show significant changes in high modulation schemes as shown in Fig. 4.

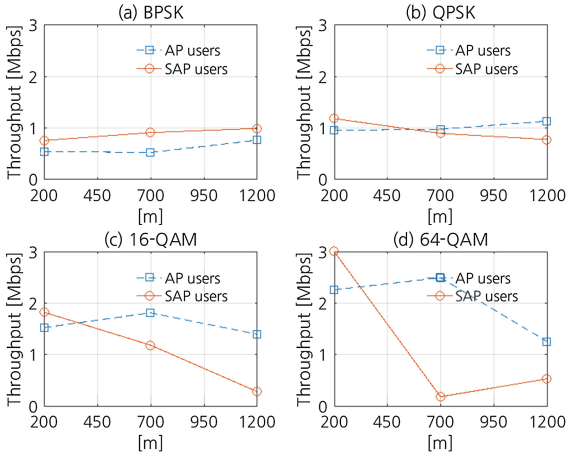


Fig. 4. Per user throughput of wireless users associated with normal-cell APs (AP users) and small-cell APs (SAP users).

For the simulations in Fig. 4, we use three different sizes of areas: 200 m × 200 m, 700 m × 700 m, and 1,200 m × 1,200 m. In each scenario, 5 normal-cell APs of 100 mW transmit power and 5 small-cell APs of 10 mW transmit power send CBR traffic of 10 Mbps during 4 s to their associated MNs from total 10 mobile nodes of 100 mW transmit power. The locations of mobile nodes and access points are randomly selected using independent Poisson point processes. For each scenario, the four modulation schemes are used to investigate the effects of the size of the areas on SINR. All simulations are repeated 5 times with different seed values.

The simulation results in Fig. 4 show that small-cell users achieve lower per user throughput than that of normal-cell users when high modulation schemes such as 16-QAM and 64-QAM are used in wider areas of 700 m × 700 m and 1,200 m × 1,200 m.

4 Conclusion

This paper investigates the effect of introducing small cells in IEEE 802.11 wireless networks. With simulations, we show the associations between mobile nodes and access points in heterogeneous networks with Friis's propagation model can be analyzed with the multiplicatively weighted metric of Voronoi diagram using the square roots of transmit powers as the weights of cells. We also show that per user throughput of small-cell users can be lower than that of normal-cell users when they use high modulation schemes.

Acknowledgements. This work was supported by 2015 Hongik University Research Fund.

References

1. Andrews, J.G.: Seven ways that HetNets are a cellular paradigm shift. IEEE Commun. Mag. **51**, 136–142 (2013)
2. NGMN (Next Generation Mobile Networks) Alliance: NGMN 5G white paper (2015)
3. The Network Simulator - ns-2. http://www.isi.edu/nsnam/ns
4. Okabe, A., Boots, B., Sugihara, K., Chiu, S.N.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. Wiley, Chichester (2000)
5. Andrews, J.G., Baccelli, F., Ganti, R.K.: A tractable approach to coverage and rate in cellular networks. IEEE Trans. Commun. **59**(11), 3122–3134 (2011)
6. Friis, H.T.: A note on a simple transmission formula. Proc. IRE **34**(5), 254–256 (1946)
7. Qi, C., Felix, S.-E., Daniel, J.: Overhaul of IEEE 802.11 modeling and simulation in NS-2. In: Proceedings of the 10th ACM MSWiM 2007, pp. 159–168 (2007)
8. Qi, C., Felix, S.-E., Daniel, J., Marc, T.-M., Luca, D., Hannes, H.: (2008). https://dsn.tm.kit.edu/medien/downloads_old/Documentation-NS-2-80211Ext-2008-02-22.pdf
9. Widmer, J.: NO Ad-hoc Routing Agent. http://icapeople.epfl.ch/widmer/uwb/ns-2/noah
10. Two-ray ground reflection model (2011). http://www.isi.edu/nsnam/ns/doc/node218.html

An Enhanced Reliable Message Transmission System Based on MQTT Protocol in IoT Environment

Hyun Cheon Hwang[1(✉)], Ji Su Park[2], Byeong Rae Lee[1], and Jin Gon Shon[1]

[1] Department of Computer Science, Graduate School, Korea National Open University, Seoul, South Korea
{panty74, brlee, jgshon}@knou.ac.kr
[2] National Center of Excellence in Software, Chungnam National University, Daejeon, South Korea
bluejisu@cnu.ac.kr

Abstract. IoT devices are widely used in everywhere for collecting data, controlling home equipment and so on. Also, MQTT protocol is used for sending a message in IoT devices because it is designed for a light-weight device and unreliable network condition. However, MQTT protocol has vulnerability to maintain order between messages against messages order is very important in some home automation such as controlling gas valve. In this paper, we design and implement a reliable message transmission system using MQTT protocol to maintain messages order.

Keywords: IoT · MQTT · Push protocol · Reliable message transmission system · Message order

1 Introduction

Network traffic consumption is being moving from desktop device to mobile device and mobile network traffic CAGR is increasing very highly. The network traffic of non-smartphone and M2M in 2015 is 89,630 TB/mo. and 99,222 TB/mo. and is expected 229,720 TB/mo. and 2,058,792 TB/mo. in 2020 according to CISCO [1]. These network traffic CAGR is each 25 % and 83 %. This means M2M system is increasing very highly in our life. The rate of increase of M2M network traffic will be 310 % in 2020 against 2015 [1].

Push protocol is widely used in IoT devices under M2M environment to transmit messages to other devices or servers. Push protocol is efficient protocol than pooling protocol because push protocol does not need to check there is a new message in a server regularly. Specially, MQTT protocol is light-weight protocol for a low-power small device [2, 3]. However, MQTT protocol only guarantees each message delivery but does not guarantee order of messages. Also, cloud IoT platforms which contains MQTT protocol such as Amazon IoT, Microsoft Azure don't support all QoS of MQTT protocol for performance issue [4, 5].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_153

Exact message delivery and guarantee of message order is mandatory in IoT security business area because wrong messages transmission can be caused an accident for real life such as wrong smoke detection, gas and water leakage detection.

Therefore, in this paper we research how messages can be transmitted exact once under right messages order.

2 Related Research

2.1 Internet of Things(IoT) and IoT Cloud Platform

The Internet of Things(IoT) is the network of physical objects or "things" embedded with electronics, software, sensors and network connectivity [6]. These objects do collect or exchange the data. For example, home-automation of LG makes people can turn off a light or lock a gas valve remotely [7].

IoT devices are designed for working alone without human decision. Some IoT Environment such as marketing messages sending so exact messages delivery is highly important. It could happen accident if there is a message missing or a wrong message delivery. People would like to use sensor devices such as smoke detectors, gas and water leakage detectors in home automation [8] and it is highly important that deliver sensor data exact once in order. If a gas valve control IoT device's message order is reversed for turning on and off, the house can fire because of wrong messages. Therefore, research of reliable message delivery and development are being going continuously.

Cloud platform is common strategy in fast changing IT market for productivity maintenance of mass data and reducing cost. Also, Korea government encourages k-cloud platform [9]. IoT is being also one of important part of could platform and several most powerful could platform supports IoT functionality. AWS IoT [4] is one of Amazon Web Service. AWS IoT supports IoT device SDK to connect a hardware device and mobile application. And that provides message communication by using MQTT protocol. Microsoft Azure IoT Hub [5] supports MQTT protocol to connect with IoT devices. However, both of IoT cloud platform only supports MQTT QoS level 0 and 1. It means a message can be transmitted at least once and a message can be transmitted more than once.

2.2 MQTT Protocol

MQTT protocol is the message push protocol released by IBM [10]. MQTT protocol was designed to transfer the messages reliably under the low-bandwidth network condition and long network delay. Currently cloud IoT platform such as Amazon AWS IoT and Microsoft Azure IoT Hub uses MQTT protocol to send a messages to devices. Facebook instant messenger application for mobile device also uses the MQTT protocol [10] as well.

MQTT protocol consists of subscriber, publish and message broker. Topic is a message string for each client and it consists of one or more levels. Each level is distinguished by slash (/) and the structure looks like tree structure. MQTT protocol

defines 3 levels of QoS (Quality of Service) which is an agreement between sender and receiver of a single message regarding the guarantees of delivery a message. There are 3 QoS levels in MQTT [3].

QoS 0: At most once. Do not guarantee a message delivery.
QoS 1: At least once. Can deliver a message more than once.
QoS 2: Exactly once. Exactly once with 4 level handshaking.

A system which needs reliable message delivery uses QoS level 2. However, QoS 2 does not guarantee messages order but only guarantee single message delivery. In addition, most of IoT cloud platform still does not support QoS 2, so some IoT device has to use MQTT protocol with QoS 1. In this situation, message delivery cannot be guaranteed when time delay has occurred because of network condition as shown picture, Fig. 1.

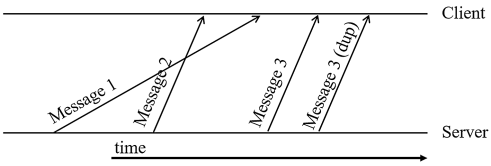


Fig. 1. Reversed or duplicated messages order in message sending flow

3 Design of a Reliable Message Transmission System

3.1 An IoT Sensor Device and an IoT Control Device

An IoT devices is classified into two types. One is an IoT sensor device and the other is an IoT control device. An IoT sensor device is the device can collect a data such as temperature or control a home equipment. An IoT control device is the device can show data or control an IoT sensor device such as turning on/off. In many cases, an IoT control device is a smart-phone. One IoT sensor device connect with more than one IoT control devices normally. The relationship between an IoT sensor device and an IoT control device is similar with 1:N, therefore, the mechanism to send a message reliably is different to both of an IoT sensor device and IoT control device.

3.2 A Expanded Message with Sequence Field

A message order and exact once delivery is needed for reliable messages delivery. So we define a message sequence field and every message can be transmitted with the message sequence field to check message order and exact once delivery. A sequence is generated and published with each message. We will expand a message field to contain a sequence. First 4 bytes of a message is reserved for sequence and sequence field is expressed as hex string, as shown in Fig. 2. Min. sequence value is *0000* and max.

/IoT_A/temperature/message1 → /IoT_A/temperature/####message1

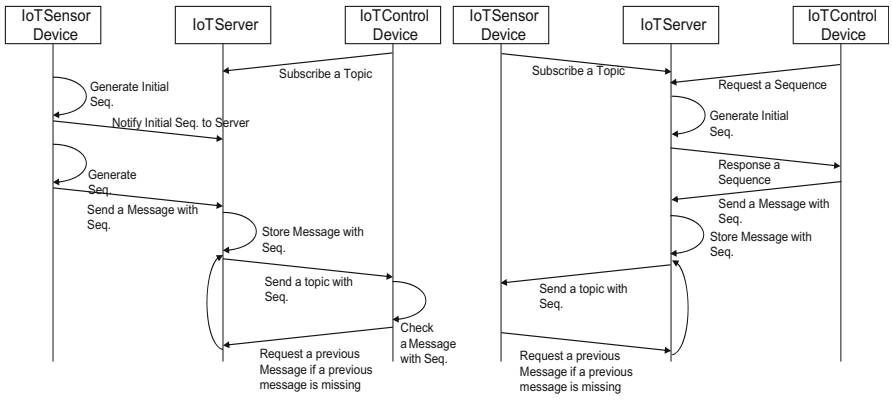
Fig. 2. Expanded messages for a sequence

sequence value is *FFFF*. It means the sequence field can handle 65,535 messages. This mechanism is better than BIT expression in other research [10] because a MQTT protocol message has to be represented by ASCII string. BIT expression can cause problem while sending a message. In addition, this sequence field uses only 4 bytes therefore, it does not affect too much network traffic.

3.3 A Reliable Message Transmission System Structure

An IoT sensor device generates initial sequence value and publish to a server because the server must know which sequence number is initial start value before publishing a message. An IoT sensor device generates a sequence number and an expanded message when publishing each message. On the other hands, An IoT control device requests a sequence number to the server before publish a message because the number of an IoT control device could be more than one and every IoT control device's sequence number have not to be duplicated. Both of an IoT sensor device and an IoT control device checks a message sequence and waits or requests a previous message if a message order is wrong when receiving a message from the other side, as shown in Fig. 3. There is a storage for a sequence number in an IoT sensor device and an IoT Server.

Figure 4 shows the reliable message transmission system structure. There is message repository to resend a missing message in an IoT server. Also, there are sequence repository in both of an IoT sensor device and an IoT server.



(a) IoT Sensor device → IoT Control device (b) IoT Control device → IoT Sensor device

Fig. 3. Data flow in reliable message transmission system

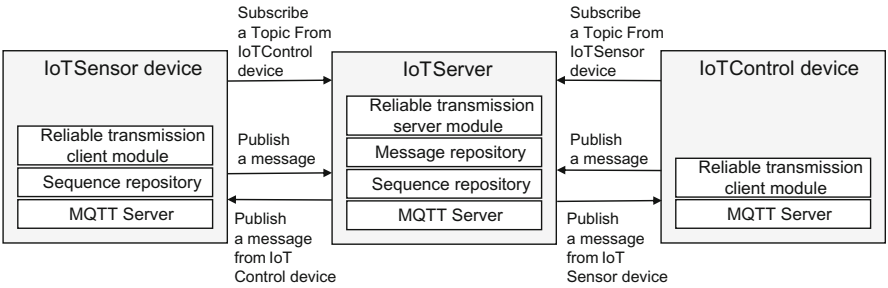


Fig. 4. Reliable message transmission system structure

4 Experiment and Analysis

4.1 Experiment Environment

We verify a reliable message transmission system described in this paper through this experiment. The experiment environment for this is in below Table 1. Also, we use the network simulator named clumsy [11] to make unstable network condition And we had experiment messages from both side of the client send without missing.

Table 1. Experiment environment

Item	IoT server	IoT senor device	IoT control device
OS	Ubuntu 14.04.3 on AWS	Ubuntu mate 16.04 on Raspberry Pi	Microsoft Windows 10
RDBMS	MySQL v5.5.50	–	–
Message broker	mosquitto v1.4.9	mosquitto v1.4.9	mosquitto v1.4.9

4.2 Result of Experiment

MQTT message broker without a reliable message transmission system did not send messages in order under unstable network condition. We made unstable network packet such as drop, lag and out of order by using the clumsy software and about 9 % messages sent in irregularly. However, the reliable message transmission system sent whole message in order without missing, as shown in Table 2.

Table 2. Experiment result

Item	MQTT protocol	Reliable message transmission system
Wrong messages rate	4 %	0 %

5 Conclusion

In this paper, we described and experimented the reliable message transmission system based on MQTT protocol to guarantee message order for IoT environment. We confirmed the reliable message transmission system can deliver messages sequentially without a long time delay. In future research, we will research more the reliable message transmission system in more complex environment.

References

1. Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2015–2020. http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.pdf. Accessed 30 July 2016
2. Kim, S., Kim, D., Oh, H., Jeon, H., Park, H.: The data collection solution based on MQTT for stable IoT platforms. J. Korea Inst. Inf. Commun. Eng. **20**, 728–738 (2016)
3. MQTT Protocol Specification. http://docs.oasis-open.org/mqtt/mqtt/v3.1.1/os/mqtt-v3.1.1-os.html. Accessed 25 July 2016
4. Amazon AWS IoT Protocols. http://docs.aws.amazon.com/iot/latest/developerguide/protocols.html. Accessed 18 July 2016
5. Microsoft Azure IoT Hub. https://azure.microsoft.com/ko-kr/documentation/articles/iot-hub-mqtt-support/. Accessed 18 July 2016
6. Internet of Things. https://en.wikipedia.org/wiki/Internet_of_things Accessed 18 July 2016
7. LG IoT@Home. http://www.uplus.co.kr/ent/iot/IotopenInfo.hpi?mid=7144. Accessed 18 July 2016
8. Kelly, S.D.T., Suryadevara, N.K., Mukhopadhyay, S.C.: Towards the implementation of IoT for environmental condition monitoring in homes. IEEE Sens. J. **13**, 3846–3853 (2013)
9. Baek, S.I., Shin, J.Y., Kim, J.W.: Exploring the Korean government policies for cloud computing service. J. Soc. e-Business Stud. **18**, 1–15 (2013). 8
10. Hwang, H.C., Park, J., Shon, J.G.: Design and implementation of a reliable message transmission system based on MQTT protocol in IoT. Wireless Pers. Commun. 1–13 (2016)
11. Clumsy. https://jagt.github.io/clumsy/. Accessed 16 July 2016

Implementation of a Smart IoT Factory Using an Agricultural Grade Sorting Device

Seokhoon Jeong, Ji Yeon Lee, Kuk Won Ko, and Sangjoon Lee(✉)

Department of Information Communication and Display Unit,
SUN MOON University-Chungnam, Asan-si, Republic of Korea
mcp94lee@sunmoon.ac.kr

Abstract. The purpose of this study is to implement the smart IoT factory through developing an automatic 6-years fresh ginseng grade classification device. The washed 6-years ginseng from farmland should be sorted 3 rating by a classification criterion such as weight and shape but this classification process has been conventionally performed manually and there are increasing sorting process costs. To overcome this disadvantage, we developed an automatic gin-seng sorting device. The 6 years ginseng put into the device for the classification is designed to perform such as the weight estimation and the shape analyzed by the image processing procedure, and classification result is sent to the factory server over the network. Evaluating the performance of developing machine experiment with 100 6-years ginseng showed a high recognition rate for 94 % for grade 1, 98 % for grade 2, and 90 % for grade 3.

Keywords: Agriculture classification · IoT factory · Image analysis

1 Introduction

Ginseng and red ginseng originate from the Republic of Korea. Red ginseng products that are manufactured and sold by the Korea Tobacco & Ginseng Corporation have been acknowledged around the world for their quality and reliability. An objective quality classification is required to grade 6-year-old fresh ginseng for the manufacture of red ginseng rather than depending on the human eye or experience [1]. This grading has caused controversy with cultivators every year. It is reported that hundreds of millions of won and 1530 skilled inspectors are required to grade raw ginseng each year. Automatic classification for some if not all 6-year-old fresh ginseng for the manufacture of red ginseng can be expected to greatly reduce the inspection costs [2, 3]. The purpose of this study was to develop an automatic 6-years ginseng sorting machine and the classification result is to send the factory server, and in the server, perform saving a daily classification result, monitoring the current classification status, and to effectively perform production management.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_154

2 Material and Method

2.1 The Design of Automatic Ginseng Sorting Machine

The Fig. 1 is showed concept of an automatic ginseng sorting machine and it is composed of the first inlet conveyor belt and the secondary classification conveyor belt. The incoming ginseng is estimated the weight by the Fig. 1-1 camera and moving to the bottom side. At this time, the ginseng is captured the 4 side shape by the 4-way camera as shown in the Fig. 1-‘2. The inletting ginseng can be sorted by estimating weight algorithm and shape analysis algorithm 1st, 2nd, and 3rd grade, respectively. Also, the sorting result is show up the LCD monitor and transmits to the factory server throughout inner network. Time that the ginseng transferred from point A to B is 0.5 m/s, and image analysis execution speed is within 0.1 s/sample. Also, the developing machine can sort 2 to 3 pieces/second.

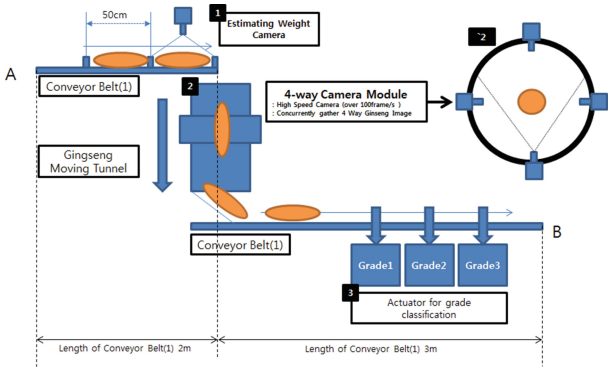


Fig. 1. The concept of the automatic ginseng sorting machine

Figure 2 shows actual developing system which can be confirmed control box for conveyor belt control, the display unit for checking image processing result, inlet and sorting conveyor blet.

Fig. 2. The actual developing ginseng sorting machine

Figure 3(a) is shows a ginseng inlet conveyor belt which is configured by black color for efficient image processing. The four sorting actuator which can control module and located is classification conveyor belt as shown the Fig. 3(b).

(a) A ginseng inlet conveyor belt

(b) A sorting conveyor belt

Fig. 3. The two step conveyor belt of an automatic ginseng sorting machine (a) A ginseng inlet conveyor belt (b) A sorting conveyor belt

2.2 The Controller of Automatic Ginseng Sorting Machine

The main control box is needed for efficient controlling of a ginseng sorting machine and the Fig. 4 shows block diagram of the control system box which is consist of a power controller, 2 AC induction motor drivers of conveyor belt, 4 the pneumatic actuator controller, a ginseng position sensor of on the conveyor and light controller for image gathering camera. The five cameras are connected to control server by USB connection as shown in Fig. 4-1. After that the finishing imaging analysis, the classification information send to the factory server and showing up a display unit.

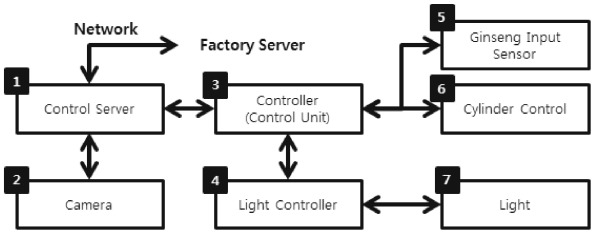
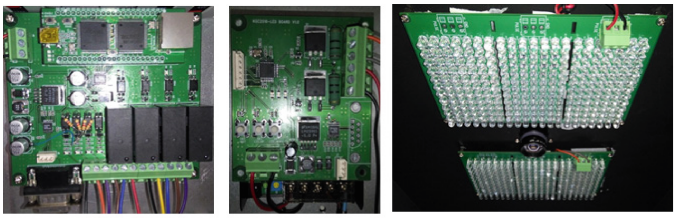


Fig.4. Control module bloc diagram

Figure 5 shows control box of a total ginseng sorting machine. The each module as shown in Fig. 4 is located inside of control box. Also, the Fig. 6(a) is snapshot of the actual developed main control PCB, the Fig. 6(b) is light controller and the Fig. 6(c) shows installing LED Array on a darkroom for a camera light source.

Fig. 5. Control box



(a) Controller (b) Light Controller (c) LED Array

Fig. 6. The actual PCB snapshot of control module (a) Controller (b) Light Controller (c) LED Array

2.3 The Estimating Ginseng Weight Algorithm

The original 1920 × 1080 pixel color image is converted to a 256 level (8 bit) or 0 (black) to 255 (white) gray level value based on the differences in light and shade to analogize the correlation of the image with the fresh ginseng weight. Afterwards, the gray image is filtered to extract the parameter most closely related to the weight information. The number of pixels ranging from 0 to 255 is counted in the converted gray image, and a histogram is developed that shows the distribution. The background region is eliminated from the fresh ginseng image, and banalization is performed to distinguish the body. The binary image has partial bright lighting in the background

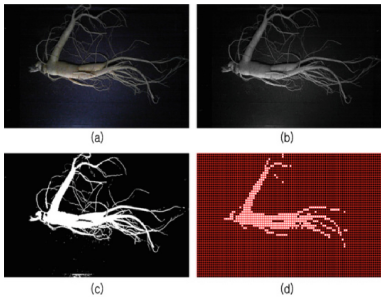


Fig. 7. Estimating ginseng weight algorithm procedure: (a) original image loading, (b) conversion to gray image, (c) image banalization, and (d) block filtering.

part besides the fresh ginseng body part. Block filtering is performed to eliminate the minute lateral roots that should have little influence on the weight of the fresh ginseng, and the remaining blocks are used to calculate the parameters. Figure 7 shows the extraction process: (a) original image loading, (b) conversion to gray image, (c) image banalization, and (d) block filtering.

A black filter divides regions of 1920 × 1080 pixel images into blocks of certain sizes. The blocks are white if the distribution of white pixels is higher than a particular threshold (%) and black if it is lower than the threshold. This filter is used to remove fine roots and show the body region that provides much of the weight.

Figure 8 shows the observed results when the length × width dimensions of the block filter were increased to 20 × 20, 40 × 40, and 60 × 60 while the threshold was fixed to 60 %. The body shape was shown while the fine roots were removed with the 20 × 20 blocks (Fig. 8a), but losses occurred due to the staircase phenomenon and resolution deterioration with increasing block size. Thus, it is important to set the proper block size and threshold.

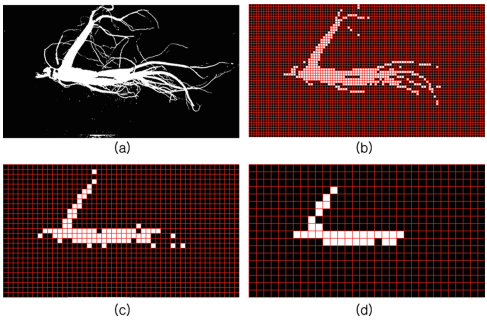


Fig. 8. The block filter result depending on a block size: (a) Binary Image, (b) 20 × 20 block size, (c) 40 × 40 block size and (d) 60 × 60 block size

2.4 The Ginseng Shape Analysis Algorithm

Just like people have different shapes and appearances, fresh ginseng specimens also have unique appearances. However, the names of certain parts are different. Figure 3 classifies the five main parts of fresh ginseng. Figure 9(a) shows the fibrous root, which is attached to the body or head region and is not present in every fresh ginseng. Figure 3(b) shows the rhizome, which is present in every fresh ginseng at the head region. Figure 3(c) shows the body of fresh ginseng, which is called the taproot. This occupies most of the fresh ginseng and is the most important part when manufacturing red ginseng. Figure 3(d) shows the leg region, which is called the lateral root; each fresh ginseng has about one to five of these roots. Figure 3(e) shows the fine roots, which absorbs nutrients in the ground.

Table 1 presents the classification criteria of grades 1, 2, and 3 ginseng provided by the KGC Ginseng Institute. Decision parameters such as the fresh ginseng weight, taproot length, ratio, and number of lateral roots were used, and the fact that experts with long

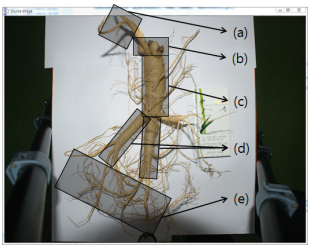


Fig. 9. Parts of the ginseng

Table 1. Grade classification criteria for 6-year-old ginseng from the KGC Ginseng Institute

	Grade 1	Grade 2	Grade 3	Off-grade
Head	Head balanced with body part	Head balanced with body part	Head balanced with body part	No limit
	Body is partially fat or curve is not severe	Body is partially fat or curve is not severe	Body is partially fat or curve is not severe	No limit
Body and leg	Body length is 6 cm or longer, more than two legs that are 5 cm or longer, but body radius must be 2/5 or less than the body length	Body length is 4 cm or longer, more than two legs that are 3 cm or longer, but body length is 8 cm or longer and is straight. The body radius must be half or less than the body length	Body length is 3 cm or longer, legs are balanced with the body, and body length of 5 cm or longer and is straight. Poor body shape with a weight of 50 g or higher is included	No limit

careers classify specimens with rough judgment without accurate measurement tools was examined. Here, (1) the height of the taproot, (2) width of the taproot, (3) height/width ratio of the taproot, and (4) width/height ratio of the taproot were selected as the grading parameters of fresh ginseng for pattern recognition training and test data.

3 Experiment and Result

3.1 The Result of Estimating Weight

Regarding the threshold in the banalization in the gray image, the block filter settings were fixed (block size: 10 × 10, threshold: 90 %) and the correlation coefficient from the linear regression analysis was observed while changing the banalization threshold between 41 and 50. A threshold of 43 was determined to have the highest correlation (Table 1). Afterwards, the length × width of the block filter and threshold in Tables 2 and 3 were fixed to the banalization threshold of 43. Based on increasing the block filter size from 11 × 11 to 20 × 20 pixels and increasing the threshold from 86 % to

Table 2. Ginseng grade recognition results

Number of training data	Grade 1 recognition rate (%)	Grade 2 recognition rate (%)	Grade 3 recognition rate (%)	Average recognition rate (%)	FRR (%)	FAR (%)
10	94	98	90	94	6	5.4
15	93.33	97.77	88.9	93.3	6.6	5.9
20	96.66	100	80	92.2	7.7	6.3

95 % in the linear regression analysis, the highest settings for the correlation coefficient were set to an 18 × 18 pixel block size and 90 % threshold.

The correlation coefficient for the number of output pixels and fresh ginseng weight had a maximum value of 0.9162 when a block filter was applied with a banalization threshold of 43, size of 18 × 18, and threshold of 90 %. The estimated regression was y = 2.517x + 45.82, and analysis of variance was performed on the estimated regression equation to test the significance as shown Fig. 10. It was considered to be significant with a p-value = 4.4652E−51 based on an F-test with a 0.05 significance level and decision coefficient of r2 = 0.839. This means that 83 % of the data is explained by the estimated regression equation.

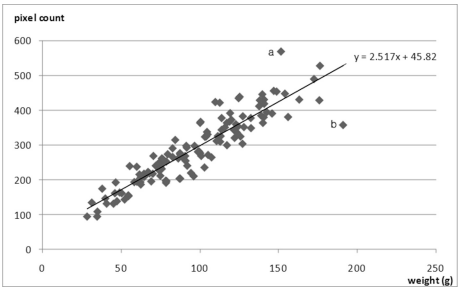


Fig. 10. The result of regression analysis.

3.2 The Result of Shape Analysis

We used SVM(support vector machine) algorithm for shape pattern classification [4, 5], Fig. 11 shows the grade classification method based on pattern recognition. The training data and test data for pattern recognition were designed so that they would not repeat. The correct match rate (%) of each grade, mean correct match rate, FRR (false rejection rate), and FAR (false acceptance rate) indices were used for performance evaluation. Table 2 presents the recognition rate and performance according to the number of training data when each of the four parameters were used. The recognition rates were 94 % for grade 1, 98 % for grade 2, and 90 % for grade 3. A high recognition performance of 6.0 % FRR and 5.4 % FAR was obtained with four parameters and 10 points of training data, as indicated in Table 3.

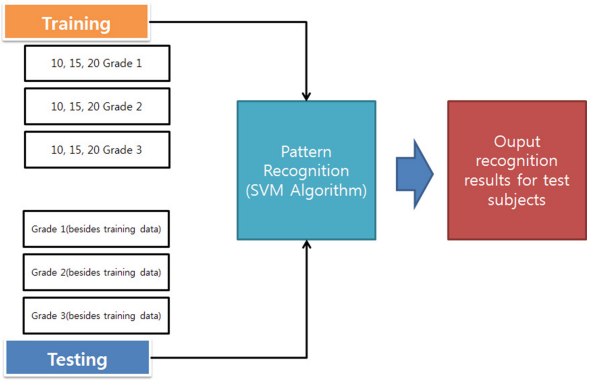


Fig. 11. The Ginseng grade classification method based on pattern recognition

4 Discussion and Conclusion

The purpose of this study was to develop an automatic 6-years ginseng sorting machine and the classification result is to send the factory server, and in the server, perform saving a daily classification result, monitoring the current classification status, and to effectively perform production management. The classification ginseng is judged by image processing procedure such as weight estimation and shape pattern analysis procedure. A ginseng, in the range of weight 75 g ~ 40 g is determined as a 2 grade, weight 40 g ~ 30 g is determined as a 3 grade or more of the weight 75, it is determined as a candidate of 1 grade. After weight estimation procedure, the 1 grade group of candidates are performed the evaluation determination by the shape analysis again to determine 1 to 3 grade. Evaluating the performance of developing machine experiment with 100 6-years ginseng showed a high recognition rate for 94 % for grade 1, 98 % for grade 2, and 90 % for grade 3.

Acknowledgement. This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea(NRF-2014H1C1A1066998) and supported by KGC ginseng research institute.

References

1. Kang, J.Y., Lee, M.G., Kim, Y.T.: Automatic decision-making on the grade of 6-year-old fresh ginseng (Panax ginseng C. A. Meyer) by and image analyzer - I. shape and weight analyses according to the grade of fresh ginseng). J. Ginseng Res. **20**(1), 65–71 (1996)
2. Jeong, C.M., Shin, J.S.: Comparison of grade of raw and red ginseng on each factor of quality in Korean and American ginseng. Korean J. Med. Crop Sci. **14**(4), 229–233 (2006)

3. Kang, J.Y., Lee, M.G.: Automatic decision-making on the grade of 6-year-old fresh ginseng (Panax ginseng C.A. Meyer) by an image analyzer 1. shape and weight analyses according to the grade of fresh ginseng. J. Ginseng Res. **26**(1), 6–9 (2002)
4. Cortes, C., Vapnik, V.: Support-vector network. Mach. Learn. **20**, 273–297 (1995)
5. Burges, C.J.C.: A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)

Segmentation and Counting of Cell in Fluorescence Microscopy Images Using Improved Chain Code Algorithm

Yeji Na[1], Sangjoon Lee[2], Jonggab Ho[1], Hwayung Jung[1], Changwon wang[1], and Se Dong Min[1(✉)]

[1] Department of Medical IT Engineering, College of Medical Science, SoonChunHyang University, 1521, 22, Soonchunhyang-ro, Asan, Chungnam 336-745, Korea
{nayejil649, hodoril988, show7kr, changwon, sedongmin}@sch.ac.kr

[2] School of Mechanical and ICT Convergence Engineering, College of Engineering, SunMoon University, 406, 221, Sunmoon-ro, Asan, Chungnam 31460, Korea
2mcp94lee@sunmoon.ac.kr

Abstract. This study aims to automatically segment of oval cell in fluorescence stained cell image and quantify cell counts. For this study, an algorithm for oval cell contour tracking was suggested based on the classic chain code method and overlapped cells were segmented using border line angle variation information. For verifying the accuracy of the suggested method, our method and Freeman's chain code method were applied to the same oval cell images. Then the border line tracking results were identified and the execution speed and computation per pixel were compared. Also, it was compared with the segmentation result of the Watershed technique, which is a general region-based segmentation, for evaluating the cell segmentation result with the naked eye. We applied an automatic algorithm to quantify cell counts in 20 cell images. For verifying the accuracy of cell counting, our algorithm was compared with the result of the manual counting method and ImageJ tool-based counting method.

Keywords: Fluorescence microsccpy image · Cell segmentation · Chain code technique · Oval cell · Cell counting

1 Introduction

Recently, many researchers were measured from cell images of the microscope for check of medicine treatment and pathologic diagnosis of cellular tissue in bioinformatics field. Cell image processing was mostly performed by a specialist's subjective visual inspection, and it was required a lot of time, costs, high concentration as well as derived proper result. To acquire an objective and high reproducibility of cell image, many studies have been conducted about automated analysis of cell images. [1–6] In case of interpretation of the cell image, image segmentation is the most basic and essential step. However, it is hard to exact segmentation because different many noise and environment.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_155

A method of image segmentation is generally used region-based [7], edge-based [8], histogram-based [9], cluster-based [10], Morphological-based [11]. In the region-based method mainly utilizes the watershed algorithm developed by Beucher [12]. Edge-based segmentation method of Freeman chain code [13] schemes introduced into 1961 is mainly used. Freeman Chain Code (FCC) Algorithm is mainly used tracking algorithm for object boundaries. Recently, the FCC used as for the cell segment method of bladder tissue cell, skin tissue cell, red blood cell [14, 15].

In this paper, we proposed suitable algorithm that edge detects of ellipse shape cell based on boundary line tracking method of chain code. In the case of each other overlapped cell to find overlapping point to track connected components of eight-connectivity mask of chain code, and segment single cell using distance based linear interpolation. Proposed segmentation algorithm is total fifteen images are applied and as result measured cell count. The results of cell count is compared with visual analysis results and cell count to get free cell analysis software ImageJ. Also, it's performance was verified compared with the general watershed algorithm used to evaluate division accuracy.

2 Method

This section explains how to segments of cell images. In this paper, we used MKN-28 a gastric cancer cell image of 70-year-old Mongolian woman with taken using a fluorescent microscope. MKN-28 cell usually grows well with a large round shape. We propose an algorithm to automatically segment of based on the oval-shaped cells. The proposed methods consist of four steps. The four main steps in the method were as follows: Gaussian filtering and binarization, contour tracing, overlapped points extraction, linear interpolation. First, it is smoothed using a 3×3 Gaussian Kernel in order to reduce the noise of the image. And then it used Otsu's method of binarization that determined the Global fixed threshold. It segment of cell and background of image using a brightness difference. Secondly, in order to obtain contour information on the cell, applying the algorithm of the improved chain code in the binarized cell image. Thirdly, extracts the overlapping points of the cluster cells through the analysis of the angular variation in the connecting element of the cell contour. Fourth, Overlapping points are connected to each other by using a linear interpolation between the points at close range.

2.1 Freeman Chain Code Algorithm

In order to track the boundaries, Freeman Chain code Algorithm uses 8-way formulated sequences. The Eight ways are represented in integer from 0 to 7 and the formula is specified in Eq. 1.

$$n' = (n + 5)\ \&\ 7. \tag{1}$$

The formula above begin at n to examines contingent pixels spinning clockwise to find a pixel value n' specifying object's boundary. When it found a boundary pixel,

it moves on to it and continue to examine whether there is the boundary around it. The Algorithm finishes the tracking process when it has completed the processing of every single point and finally get back to the first pixel.

2.2 Proposed Chain Code Algorithm

The previous Freeman's Chain code Algorithm is a strong method of tracking boundary, but the round cell image of this study doesn't have great angular variation on which the pixel is linked to the next pixel. Therefore, it tends to examine the unnecessary pixels that have low possibilities of becoming boundary. It makes the process much slower because the computation increases. So this study suggests refined Chain code Algorithm for the boundary tracking of round-form cell image. The direction vector of chain code in the proposed method is as Fig. 1. In this method, the movement from one pixel to the next pixel requires examination of the contingent 8 pixels from the current pixel. Preferentially it examines the previous direction it moved from. It's recommended the soft-curve boundary doesn't make any great angular variation because of its ellipse form. When the pixel value in the search direction doesn't mean the boundary of an object, it proceeds −45 degree pixel examination in clockwise direction. When the pixel value means the boundary of an object, it proceeds +45 degree pixel examination in counter clockwise direction. The processing order for this algorithm is represented in the box.

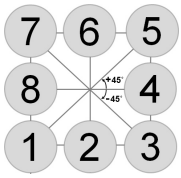


Fig. 1. The direction vector of the proposed chain code

Step 1. Find pixels on which the boundary of an object begins as you scan from top to bottom and from left to right of the input image respectively.
Step 2. Determine whether there is any boundary pixels as you scan 8 pixels contingent to the current pixel. Flag value is set to 2. (Note: When the current pixel is the beginning point of the boundary, the early search direction is set to east(ip=4). Otherwise, it searches through the direction it moved from.)
Step 3. In order to move from on pixel to another, select one from the following two options. Also, our formula is "n'= n + d"

(1) When there is a boundary pixel in the search direction:

- If the flag value is 0, it moves toward the specified direction. (proceed to the Second Step)
- Otherwise, It examines the +45 degree direction pixel as it moves from the current search direction(n) to the next search direction(n'). Then, the flag value is set to 1.

(2) When there is no boundary pixel in the search direction:

- If the flag value is 1, it moves toward the −45 degree direction pixel. (proceed to the Second Step)
- Otherwise, It examines the −45 degree direction pixel as it moves from the current search direction(n) to the next search direction(n'). Then, the flag value is set to 0.

Step 4. The search is completed when the current pixel location is equal to the beginning point of the boundary.

When the boundary tracking is done according to the specified stages above, the process is much faster as the computation is much less than the previous Chain code Algorithm. The following is a table in which comparison was made between the previous Freeman Chin code Algorithm and the new Algorithm proposed in this study. As we can see in Table 1, the result of the boundary tracking using our new method is 100 % same with the result using Freeman Chain code Algorithm. However, the method used in this study proceeded 138 turns less searching than the previous one. The new method also was 0.331 s faster when tracking the boundary.

Table 1. The result of tracing the contour using FCC and our new method.

	Our new method	FCC method
Contour chain code	4 4 3 4 3 3 2 2 2 2 2 2 2 2 2 2 1 2 1 2 8 8 1 8 8 7 8 7 6 6 7 6 7 6 6 6 6 5 6 6 5 4 5 4 4	
Amount of computation	94	252
Execution speed	0.147 s	0.478 s

Sometimes, there are cases when the cells are overlapped as in Fig. 2(a). In order to resolve this problem, we used the border line angle variation information. The points on which the images overlap show great angular variation. Therefore, we used the angular variation information acquired by our contour tracing method above to capture the overlapping points.

We can see the curve is falling and then rising in Fig. 2(b). So we can determine the points with great direction change as the overlapping points. After determining of the overlapping points, the most nearest two points are connected using linear interpolation.

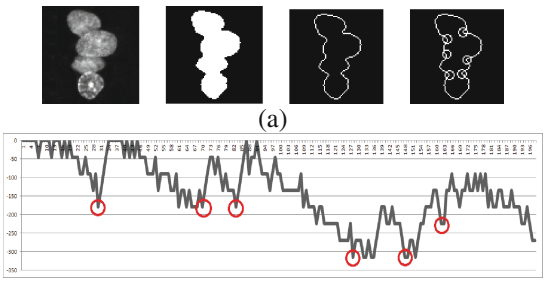


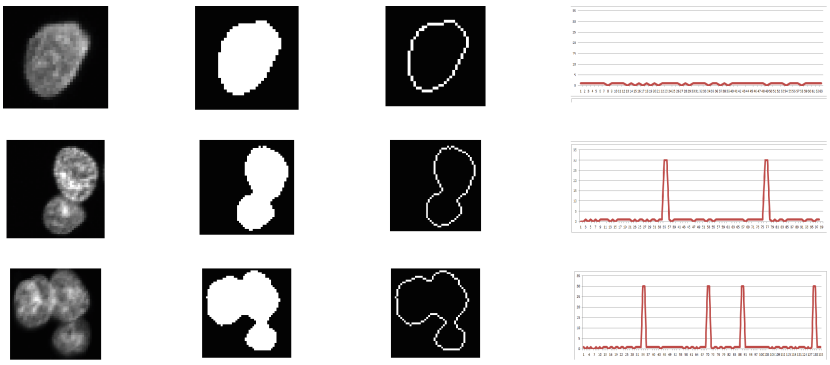
Fig. 2. (a) The case of overlapped cells and (b) The method for finding overlapped points

Use the formula (2) if you want the straight line calculated lining up two points when using the linear interpolation.

$$p(x) = \frac{y_i(x - x_{i+1}) - y_{i+1}(x - x_i)}{x_i - x_{i+1}} \tag{2}$$

3 Result

To trace the boundaries of the proposed scheme and to split the cell, We are in the windows 7 Operating System Using the NI LabWindows CVI 2013 tool was implemented algorithm in C language. In the case of overlapped cells and single cell in the microscope image were exhibited under the Fig. 3(a–c). The overlapped cells point obtained by using the angular variation information, it was exhibited the peak value from the graph of Fig. 3(d).



a. original image b. binary image c. contour image d. Peak of overlapping cells

Fig. 3. The case of overlapped Cell image and graph

Under the Fig. 4(a) is part of the cell image, the Fig. 4(b) is binary image using the Otsu binarization, the Fig. 4(c) is the result of contour tracing using the proposed in this paper. Also Fig. 4(d) is segmented cell using the proposed method. The Fig. 4(e) is exhibited result of cell segmentation using the Watershed Algorithm. The proposed method shows no significant difference when compared with the result of Watershed Algorithm by naked eyes. However, we find that the area of the cell is different to two results. Compared with the watershed algorithm, A loss of the cell area information on images was decreased from our algorithm.

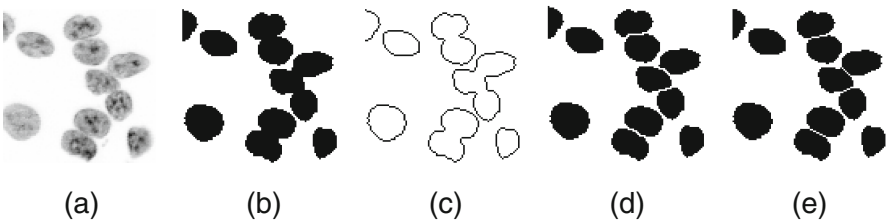


Fig. 4. Comparative results of the watershed method and our method

The following Table 2 describes measured cell count using the visual inspection and purposed method in this paper (A), Image J (NIH, National Institute of Health) free image analysis software (B) from the 15 pictures.

Table 2. The result of cell counting using two methods

Test image number	Manual cell count	Auto cell count (A)	ImageJ cell count (B)
1	64	64	64
2	67	67	64
3	69	65	68
4	78	78	73
5	88	85	82
6	97	89	73
7	106	99	101
8	135	132	126
9	139	139	134
10	151	150	151
11	154	154	150
12	157	155	140
13	159	148	145
14	162	159	148
15	178	175	173
Average accuracy (%)		**97.3**	**93.8**

When measured the segmentation accuracy which compared A and B method based on visual inspection, accuracy of A method showed at 97.3 % and B method showed at 93.8 %. Also, the error rate of Automatic Cell counts showed at 2.54 %, and the error rate of ImageJ cell counts showed 5.46 %.

4 Conclusion

In this paper, we were tracking the contour of the oval cell by applying the improved chain code algorithm. And then, it found the cell overlapped points to segments of overlapped cell analysing the angular variation information of cell edges. Overlapping or touching cell's overlapped points used to segments by linear interpolation. This method is faster than contour tracing time of Freeman Chain code about 0.33 s, also amount of computation is less than about 2.7 times. Moreover loss of cell area information was that the method was less than using the watershed algorithm. It was calculated based on the time that is applied to one cell. In fact one image, the average 120 or more cells were present. Therefore, It is expected more bigger differences that total performance speed and amount of computation when applied to the hundreds of images. The proposed algorithms have been experimentally total 15 images, it was higher than ImageJ that was free cell image analysis tool accuracy of cell coefficient 3.5 %, error rate was lower by 2.92 %. In this way, the performance of the proposed method was verified objectively more excellent. However, if the distance between overlapping points is equal, the task of cell segmentation was difficult. Also, there is a problem when the pixel of overlapping cells was only one. By supplementing the algorithm in the future if those problems are solved, it can be applied to analysis software that required cell segmentation quickly and correctly.

Acknowledgements. This work was supported by the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1066998).

This research was also supported by the MSIP (Ministry of Science, ICT & Future Planning), Korea, under the C-ITRC (Convergence Information Technology Research Center) support program (NIPA-2016-H8601-16-1009) supervised by the NIPA(National IT Industry Promotion Agency).

References

1. Buggenthin, F., et al.: An automatic method for robust and fast cell detection in bright field images from high-throughput microscopy. BMC Bioinform. **14**(1), 297 (2013)
2. Čibej, U., et al.: Automatic adaptation of filter sequences for cell counting. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE (2015)
3. Kothari, S., Chaudry, Q., Wang, M.D.: Automated cell counting and cluster segmentation using concavity detection and ellipse fitting techniques. In: 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro. IEEE (2009)

4. Wang, M., et al.: Novel cell segmentation and online SVM for cell cycle phase identification in automated microscopy. Bioinformatics **24**(1), 94–101 (2008)
5. Chittajallu, D.R., et al.: In vivo cell-cycle profiling in xenograft tumors by quantitative intravital microscopy. Nat. Methods **12**(6), 577–585 (2015)
6. Cosio, F.A., et al.: Automatic counting of immunocytochemically stained cells. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1. IEEE (2003)
7. Cheng, J., Rajapakse, J.C.: Segmentation of clustered nuclei with shape markers and marking function. IEEE Trans. Biomed. Eng. **56**(3), 741–748 (2009)
8. Wani, M.A., Batchelor, B.G.: Edge-region-based segmentation of range images. IEEE Trans. Pattern Anal. Mach. Intell. **16**(3), 314–319 (1994)
9. Wu, H.-S., Berba, J., Gil, J.: Iterative thresholding for segmentation of cells from noisy images. J. Microsc. **197**(3), 296–304 (2000)
10. Ortiz De Solórzano, C., et al.: Segmentation of confocal microscope images of cell nuclei in thick tissue sections. J. Microsc. **193**(3), 212–226 (1999)
11. Eun, S.-j., Whangbo, T.-K.: Image segmentation algorithm based on geometric information of circular shape object. J. Internet Comput. Serv. **10**(6), 99–111 (2009)
12. Beucher, S., Lantuéjoul, C.: Use of watersheds in contour detection (1979)
13. Freeman, H.: On the encoding of arbitrary geometric configurations. IRE Trans. Electron. Comput. **2**, 260–268 (1961)
14. Korde, V.R., et al.: Automatic segmentation of cell nuclei in bladder and skin tissue for karyometric analysis. In: European Conference on Biomedical Optics. International Society for Optics and Photonics (2007)
15. Chen, H.-M., Tsao, Y.-T., Tsai, S.-N.: Automatic image segmentation and classification based on direction texton technique for hemolytic anemia in thin blood smears. Mach. Vis. Appl. **25**(2), 501–510 (2014)

Empirical Study of the IoT-Learning for Obese Patients that Require Personal Training

Seul-Ah Shin[1(✉)], Nam-Yong Lee[2], and Jin-Ho Park[2]

[1] Department of Computer Science, Graduate School, Soongsil University, Seoul, Korea
sashin@ssu.ac.kr
[2] School of Software, Soongsil University, Seoul, Korea
{nylee, j.park}@ssu.ac.kr

Abstract. Modern people are spending a lot of time and money for weight management, among them an increasing number of people to the personal training for the balance of his body. However, the public does not have the expertise are not easy to analyze the state of his body to find exercises in your body. So look for a lot of people to a personal training fitness center. However, personal training is difficult to continue to exercise because lifting is expensive. Therefore, we define a new concept of IoT-learning by integrating the IoT technologies and U-learning in this paper. We also propose a personalized personal training using IoT-learning. And it was mainly practices that can be applied.

Keywords: IoT-Learning · U-health · U-learning

1 Introduction

According to the OECD report, obesity rates from the 1980 s up to now have been increased to two to three times. More than 50 % of the population was found to be overweight in more than half of the country [4]. Obesity is not simply to say that going to a lot of body weight. Obesity is a number of body fats is more state than usual. Because obesity is to induce various diseases, to a method of preventing obesity, diet, there is a need for exercise therapy. To improve the eating habits in accordance with the state of the body, to analyze the amount of body fat of each person's body, it is important to design an exercise that suits you. However, to find a person to fit the movement there is a limit. For this reason, people look for a fitness center. However, the fitness center of the user, select the exercise equipment that is not right for you, it is possible to be injured by a movement away the incorrect position. And in order to learn effectively exercise method is requires a lot of time. The use of personal training (PT) in order to solve this problem. However, a problem has occurred in the expensive cost. In addition, there is a problem that is difficult to cultivate the habit of exercise in. When trying to solve these problems, in this paper, it is possible to take advantage of the IoT-learning, recommends an exercise that suits you, see the recommendation movement smart phone, in a display such as a smart TV. Also, anytime, anywhere it is possible to provide a personalized motion. To take advantage of the IoT, me with the

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_156

motion recognition sensor on the body, while the recommended exercise, it is possible to efficiently exercise correct a disproportionate attitude to IoT sensor. Thus, relying on expensive costs and trainer of personal training can solve the part that was hard to his movement, to correct the distorted posture, it would be able to exercise, manage your body weight to practice alone it is possible to.

2 Related Researches

2.1 U-Health

Which was fused the U-health and health care and IT, specifically physicians time, a remote medical care system for medical care of the patient without the space constraints [13]. Using mobile IT equipment, at any time, it is everywhere it is possible to receive medical treatment of health care workers' technology. In addition, U-healthcare is closed, using the wireless network information and communications technology, anywhere, at any time, the evaluation of real-time health status of the individual, diagnosis, and refers to all of the services for the treatment [13]. In the user's point of view, the treatment to target the healthy people and the field of health management for the purpose of, disease prevention and physical strength is the goal.

2.2 U-Learning

Beyond the e-learning that is an online education simply based on internet experience, U-learning means the learning system that is possible anytime, anywhere, and to anyone in convenient manners [10]. If e-learning is an internet-based online learning, U-learning is a crystal of e-learning, which is the online learning in the U-learning environment with the availability of accessing the knowledge and information without the constraint of time and space via Any Network, Any Device [11]. With U-learning, it is possible to make use of everywhere in the world as a learning space being freed from the physical limitation of classroom by taking advantage of wireless Internet, Augmented Reality and Web reality technology including the things that exist in everyday life, the things that exist in the learning activity space, as well as the sensors, chips, and labels, and it is also possible to have personalized and customized learning according to the student interest, preference, learning styles, and learning contexts in intelligent learning environment [11].

3 IoT-Learning

3.1 IoT-Learning

(1) Definition
If the concept of IoT-Internet is that each thing with a built-in sensor makes a new value by being connected to each other, IoT-learning means to provide users with the training from something that made a new value. It menas not only to provide

users with customized information by transmitting data over the network and by receiving the information to correct and analyze it, but also to provide the users with training with the information that matches the customized information. IoT-learning means to provide users (learners) with the information by collecting the information via sensors if there is education needed by the users any time any where. The biggest goal of this learning is to give customized training to the users with the data obtained via sensor technology of IoT. Figure 1 shows the paradigm shifts of learning from e-learning to U-learning, then to smart-learning, and expects that the learning system in the future will shift to the learning system utilizing IoT.

(2) IoT-learning applied in rehabilitation
Enter the postures the patients receiving rehabilitation care must take heed into the devices such as smartphone. Give alarms to the users that it is not good to maintain their present posture any more by detecting the posture with sensors. If the alarm lasted for a few times, provide the users with a training that may help the back rehabilitation while watching the videos to be played on the display screen such as smart TV, laptop, smartphone which the users can watch right away.

(3) IoT-learning applying Healthy Care
This is the chair using the Healthy Care, it provides users the stretching they need by looking for the areas of body with, for example, 'shoulder tightness' which need stretching via sensors and by analyzing the collected data anomalies by making use of the sensors attached on the chair that confirms the state of the body and can find the abnormalities of the body. The customized stretching can be offered to users by the transmission of data to the display screens in close proximity of the users such as TV or smart phone.

(4) IoT-learning applied to Personal Training (PT)
The user can perform exercise by view video by searching yourself movement towards portion to be exercise in the website. And, as another method is to collect the user of health data onto the IoT, health data of the collected user dry obesity to diagnose on the basis of the BMI, of Normal, obesity I, obesity II, to obesity III divided into us to recommend the combined exercise program to the user. User prior to the recommended exercise program, attaching a sensor to the body muscles being used. Attach sensor, while watching the video, inform the user is

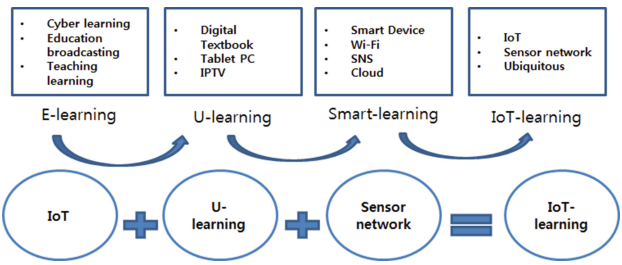


Fig. 1. IoT-learning

Table 1. BMI standards

	Normal	Obesity I	Obesity II	Obesity III
Male	Less than 20	Less than 25	Less than 30	More than 30
Female	Less than 25	Less than 30	Less than 35	More than 30

Table 2. Application fields of IoT-learning

Field of Application	Application Methods	Application Display
IoT-learning for rehabilitation	(1) Enter the postures for the rehab patients to take heed	Smartphone Tablet PC Laptop Smart appliances Wearable Smart Car
	(2) Detect the postures to take heed by sensors and give alarms to the patients if the same postures last	
	(3) Play the stretching video on the display screen in order for the patients to stretch when the alarm lasted a few times	
IoT learning utilizing healthy care	(1) A user sit on the chair with sensors attached which can find abnormalities in the body	
	(2) Once the user sits on the chair, the sensor finds where stretching is needed such as 'shoulder tightness' by analyzing the user's body and inform the user	
	(3) Even if a user does not manually search for the video associated with stretching, offer a personalized video the user requires automatically	
IoT-learning of PT (personal training)	(1) Users can proceed with exercise they want to do watching videos	
	(2) Sensors inform the users of wrong postures while they are watching the videos	
	(3) It helps the users to work out with right postures	

the wrong attitude by vibration. Users, it is possible to correct the wrong attitude to recognize the wrong attitude. The user can increase the effect of exercise (Tables 1 and 2).

4 IoT-Learning of Personal Training (PT)

4.1 IoT-Learning of Personal Training (PT)

Personal trainer after receiving a diagnosis to collect data, such as risk factors, physical fitness, body shape, physical activity from IoT sensor, helping the prescription and rehabilitation. To take advantage of the learning system so that you can perform user anytime, anywhere on the basis of simple, it is possible to make the training that

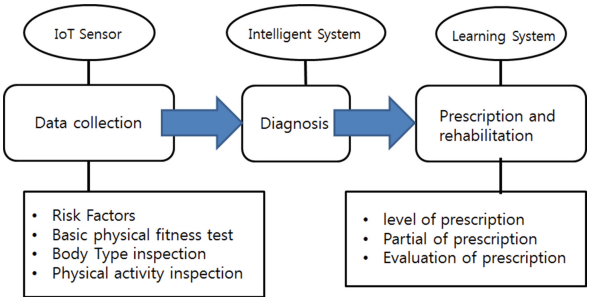


Fig. 2. IoT-learning of personal training (PT)

matches the user divided level of prescription, partial of prescription, in the Evaluation of prescription (Fig. 2).

4.2 IoT-Learning of Personal Training (PT)

(1) IoT learning application of the Personal Training case A
On the basis of the health data onto the user that IoT sensor has been collected, can we recommend an exercise program that suits the user. The user, the movement has received a recommendation program. The wrong posture while you exercise, using a sensor, induced to the calibration to the user. While advancing motion while correcting the posture that the user is repeated, the user can effectively exercise.

(2) IoT learning application of the Personal Training case B
The user selects one of the motions of the moving image that is shared among a Web site. Through a program video of a selected motion, converted in accordance with the IoT learning, it gives me reporting the muscles used in the exercise in the user, the user is attached to only the sensor. While watching the movie, users can exercise, by using a motion recognition technology, giving an alarm when the user takes an unauthorized position. The user, while the movement towards the right position, it is possible to adjust the balance of the body (Tables 3 and 4).

4.3 IoT-Learning of Personal Training Case Examples

Those 35 people that had experienced a personal training studied the use of IoT learning targets. Exercise can be professionally receiving a personal training and training has the advantage that the wrong posture correction. For that reason a higher satisfaction with personal training. However, the value was difficult because of the constant movement expensive. It is expensive because of personal training and finds videos from video sites, Web sites that training home less pay the cost aspects. But it cannot be corrected by exercise in attitude disturbance. So it answered that is difficult to continue the exercise. The advantage of the IoT learning in order to solve these

Table 3. IoT-learning of Personal Training (PT)

	Method of applying	
IoT-learning of PT (personal training)	A	(1) A Using data IoT sensor collects, recommending the combined motion to the user
		(2) On the recommendation program, the user movement
		(3) The attitude disturbed while advancing the movement informs the user at the sensor
		(4) The user, the posture corrected in the information received from the sensor, to continue the exercise
	B	(1) Users look for the motion of the video on the Web site
		(2) To convert the program to fit the video of the selected movement away IoT- learning
		(3) Attaching a sensor to the user's body before to see the video
		(4) While watching the movie, users can exercise, will help to be able to correct the incorrect position the user through the motion recognition

Table 4. Exercise program

	Exercise period (Week)	Motion duration (Minutes)	Exercise performance order	Explanation of exercise
Obesity I	1–5 W	10	1. Warm-up	Stretching
		20	2. Exercise of muscle	Table 5. Partial strength training
		20	3. Aerobic Exercise	Running Machine/Stationary Bicycle
		10	4. Stretching	Flexibility Exercise
Obesity II	6–12 W	10	1. Warm-up	Stretching
		20	2. Exercise of muscle	Table 5. Partial strength training
		40	3. Aerobic Exercise	Running Machine/Stationary Bicycle
		10	4. Stretching	Flexibility Exercise
Obesity III	12–15 W	10	1. Warm-up	Stretching
		10	2. Exercise of muscle	Table 5. Partial strength training
		60	3. Aerobic Exercise	Running Machine/ Stationary Bicycle
		10	4. Stretching	Flexibility Exercise

problems is needed. The user can reduce the burden of the cost aspect to utilize home training; it is possible to increase the satisfaction with the posture correction to the movement away.

Table 5. Partial strength training

Body			Kind of strength training	Level
Chest Exercise (CE)		CE1	Bench Press - Barbell, Flat	Level 2
		CE2	Fly - Dumbbell, Flat	Level 1
		CE3	Bench Press - Barbell. Incline	Level 2
Abdominis Exercise (AE)		AE1	Crunch	Level 1
		AE2	Leg Raise	Level 1
		AE3	V-up	Level 3
Latissimus Dorsi Muscle Exercise (LDME)		LDME1	Lat Pull Down-Machine	Level 1
		LDME2	Row Dumbell, One Arm,	Level 1
		LDME3	Pull-up - Assisted	Level 1
Thigh	Quadriceps Femoris Exercise (QFE)	QFE1	Squat-Barbell	Level 3
		QFE2	Leg Press	Level 1
		QFE3	Lunge-Barbell	Level 3
	Hamstrings Exercise (HE)	HE1	Leg Curl - Standing	Level 1
		HE2	Leg Curl - Lying	Level 1
		HE3	Multi-Hip	Level 1

5 Conclusion

In this paper, the use of IoT learning is a concept that combined the IoT sensing technology and learning system. If the user me to recommend the necessary exercise, can you explain where you put the sensor using the motion recognition technology, which is provided for the user. And provides for scheme for a sensor attached to the body can improve the efficiency of correct movement user posture. By utilizing IoT learning, can solve expensive cost issues, it is possible to set a target exercise to suit his own. Users can help weight loss. This paper presents the use scheme that can be utilized to define the concept of IoT learning was analyzed for applying draft.

References

1. Choi, W., Kim, T., Lim, C.: A real-time motion recognition algorition for a rehabilitation service. J. Korea Multimed. Soc. **10**(9), 1143–1152 (2007)
2. Yeo, N.H., Park, I.B.: Development of remodeling exercise prescription program for treatment of obesity and osteoporosis in older women. Exerc. Sci. **13**(3), 351–366 (2004)
3. Lim, B.-K., Zhang, B.-T.: Personalized menu recommendation algorithm using hypernetwork. Korea Comput. Conf. **39**(1(B)), 393–395 (2012)
4. Kim, J.-H., Park, J.-S., Jung, E.-Y., Park, D.-K., Lee, Y.-H.: A diet prescription system for U-Healthcare personalized services. J. Korea Contents Assoc. **10**(2), 111–119 (2010)
5. Moon, J.-W.: Effect of aerobic and resistance exercise to hyperlipidemia. Health Sport Med. Official J. KACEP, **8**(2) (2006)
6. Lee, J.-L., Yoon, J.: A Study on the limitation of Nintendo Wii using physical interactive interface. J. Korea Game Soc. **11**, 93–104 (2011)
7. Mcgonigal, J.: The Gamification Summit, Day1 Conference Sessions - Make it GAMEFUL (2011)
8. Essential Facts about the computer and video game industry, ESA (Entertainment Software Association) (2011)
9. McGonigal, J.: Reality is Broken: Why Games Make us Better and How They Can Change the World. Penguin, New York (2011)
10. New Media Consortium. The NMC Horizon Report: 2011 K-12th edn. (2011). emqusa.com
11. Lee, H., Kim, M., Bang, H.: Internet of things technology trends and future direction. Inf. Process. Soc. 21(2), 2014
12. Connecting Lab, Huge Connection beyond Cloud and Big Data, IoT, Window to the Future (2014)
13. Cho, H., Koh, D.: Policy implications for vitalizing U-health services based on the consumer survey and the expert interview. J. Korea Technol. Innov. Soc. **14**(3), 488–515 (2011)

Detection of Optimal Activity Recognition Algorithm for Elderly Using Smartphone

Changwon Wang[1(✉)], Sangjoon Lee[2], Jonggab Ho[1], Yeji Na[1], and Se Dong Min[1(✉)]

[1] 1521, Department of Medical IT Engineering, College of Medical Science, SoonChunHyang University, 22, Soonchunhyang-ro, Asan Chungnam 336-745, Korea
{changwon, hodoril988, nayejil649, sedongmin}@sch.ac.kr

[2] 406, School of Mechanical and ICT Convergence Engineering, College of Engineering, SoonChunHyang University, 221, Sunmoon-ro, Asan, Cungnam 31460, Korea
mcp94lee@sunmoon.ac.kr

Abstract. This is a preliminary study regarding the design of a fall prediction system. The purpose of this work is to determine an optimal classification algorithm for a fall prediction system that is able to recognize the gait difference between healthy and elderly people using a triaxial accelerometer sensor on a smartphone. To evaluate our approach, 19 people participated in our experiments. We collected accelerometer data as they performed daily activities such as walking, hobbling, and sticking, and features including the mean, standard deviation, and horizontal and vertical components were calculated. A Naïve Bayes classifier, a Bayesian network, a support vector machine, the k-nearest neighbors (k-NN) algorithm, a decision tree, multilayer perception, and logistic regression were used to classify these features using the Weka assessment tool. An 10-fold cross-validation method was carried out to classify daily activities and to compare the accuracy of the classification of daily activities for healthy and elderly people. As a result, the overall accuracy of recognition was 97.4 % for healthy adults and 71.1 % for elderly people, and the k-NN algorithm was higher than the other classification algorithms with accuracies of 99.5 % and 81.4 %.

Keywords: Activity recognition · k-NN · A triaxial accelerometer sensor

1 Introduction

Recently, falling has become a serious problem among the elderly people. It is a leading cause of death and hospitalization in elderly people and the main cause of death in people over the age of 65 [1]. According to previous studies, 30 % and 40 % of the elderly over the ages of 65 and 80, respectively, experienced a fall at least once per year. The death rate for the elderly due to falling is greater than 60 %, and half of the elderly who were hospitalized owing to falling ended up dead within a year [2]. In Korea, older adults over the age of 65 have experienced falls [3].

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_157

Owing to this, many studies related to falling have been conducted, which have utilized a triaxial accelerometer sensor and implemented fall detection systems using this sensor [4–6]. However, it is not important for the elderly that the system can detect a fall, as people have already suffered an injury since detection occurred after the fall. Thus, the prediction and prevention of a fall are very important. However, is it very difficult to implement such a system to accommodate these needs, and the related research is also insufficient. Therefore, this study aims to find an optimal algorithm to classify daily activities such as walking, hobbling, and sticking by evaluating data gathered from healthy and elderly people so that the most suitable algorithm can be suggested for implementing a fall prediction and prevention system.

This paper is organized as follows. Section 2 describes the characteristics of the subjects, the data acquisition, the data preprocessing, and the data recognition. Section 3 describes the experimental results. Section 4 summarizes our conclusions and discusses the plans for future work.

2 Method

2.1 Characteristic of the Subjects

In this study, nine elderly people selected as the subjects, which include six males and three females. In addition, 10 healthy adults, including eight males and two females, were selected. For the elderly, the age distribution was 67–94 with a mean age of 79.7 years. For the healthy adults, the age distribution was 23–28 with a mean age of 25.3 years.

2.2 Data Acquisition

Data were measured from the triaxial accelerometer of a smartphone as follows:

$$\begin{aligned} X &= (x_1, x_2, x_3, \ldots), \quad Y = (y_1, y_2, y_3, \ldots) \\ Z &= (z_1, z_2, z_3, \ldots) \end{aligned} \tag{1}$$

In addition, the corresponding stream vector magnitude was defined as follows:

$$M = (m_1, m_2, m_3, \ldots)$$

$$\text{where,} \quad m_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{2}$$

The stream vector magnitude was derived by combining the three-axis measurement data obtained by the experiment using (2). In order to acquire the triaxial accelerometer sensor data related to the gait, we adopted the data in stored the smartphone, as illustrated in Fig. 1. Data were sampled at 30 Hz and captured for 1 min. The accelerometer sensor was located at the waist, as shown in Fig. 2.

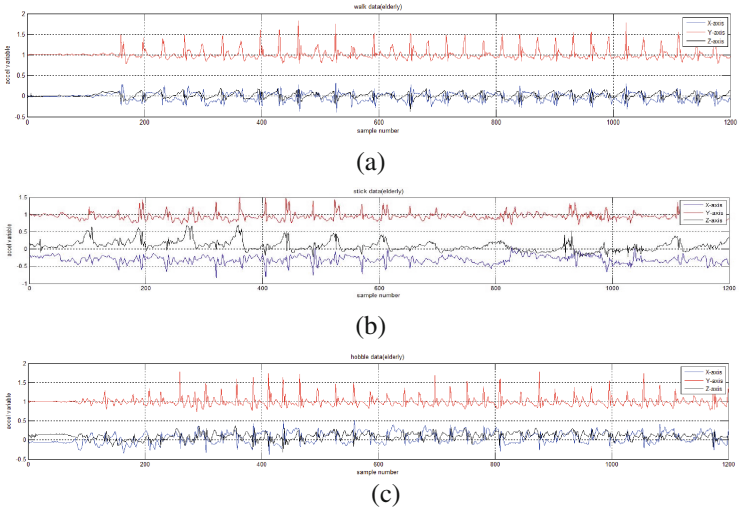


Fig. 1. Accelerometer data from elderly subjects: (a) walking, (b) sticking, and (c) hobbling

Fig. 2. Subjects and accelerometer sensor data of a smartphone

2.3 Data Preprocessing

In order to remove noise and smoothen the accelerometer data, data preprocessing was performed using moving-average filtering with five averaging points.

Owing to the continuous movement, it is difficult to obtain accurate information from the three-axis acceleration data from a smartphone. Therefore, we acquired the vertical and horizontal components for additional specification as shown in Fig. 3. $\overline{a_i}$ is the average of all measured values for each axis over the sampling interval. In the moving-average filter, L was 10. $\overline{v_{norm}}$ is the normalization of $\overline{v_i}$.

Fig. 3. Vertical and horizontal features

$$\overline{a_i} = (x_i'', y_i'', z_i''), i = 1, 2, 3, \ldots, N. \tag{4}$$

The vertical component $\overline{v_i}$ is found by applying $\overline{a_i}$ to $\overline{v_{norm}}$ as

$$\overline{v_i} = \left(\frac{\overline{a_i} \cdot \overline{v_{norm}}}{|\overline{v_{norm}}|^2} \right) \overline{v_{norm}}. \tag{5}$$

The horizontal component $\overline{h_i}$ is calculated by difference of $\overline{a_i}$ and $\overline{v_i}$:

$$\overline{h_i} = \overline{a_i} - \overline{v_i}. \tag{6}$$

2.4 Data Recognition

The accelerometer data were used as features for gait recognition. We used a decision tree, a support vector machine (SVM), a Naïve Bayes classifier, a Bayesian network, multilayer perception, logistic regression, and the k-nearest neighbors (k-NN) algorithm to classify the data using the Weka assessment tool. On this basis, we evaluated the differences in gait recognition between healthy adults and elderly people using 10-fold cross-validation for each of the data-mining techniques.

3 Results

Tables 1 and 2 summarize the accuracy for each classifier for the three types of gait patterns using a 10-fold cross validation method. For the accuracy of each classifier for healthy adults, all of the classification algorithms were greater than 95 %. Further, for elderly people, the average accuracy about all of the classification algorithms was 71 %. The precision for healthy adults for all gait patterns was greater than that for elderly people by 24 %.

In case of elderly people, the average precision for all gait patterns was 0.711. Among them, walking has a higher average precision than the other gait patterns. the average precision of each gait pattern was 0.829 for walking, 0.632 for hobbling, and

Table 1. Detailed accuracy for each classifier for healthy adults

Classifier	TP rate	FP rate	Precision	Recall	F-measure
NB	95.2 %	2.5 %	95.2 %	95.2 %	95.2 %
BN	96 %	2 %	96.1 %	96 %	96 %
SVM	97.4 %	1.2 %	97.4 %	97.4 %	97.4 %
k-NN	99.5 %	0.3 %	99.5 %	99.5 %	99.5 %
DT	97.3 %	1.4 %	97.3 %	97.3 %	97.3 %
MLP	98.8 %	0.6 %	98.8 %	98.8 %	98.8 %
Logistic	97.6 %	1.2 %	97.6 %	97.6 %	97.6 %

Table 2. Detailed accuracy for each classifier for elderly people

Classifier	TP rate	FP rate	Precision	Recall	F-measure
NB	58 %	17.2 %	68.6 %	58 %	57.9 %
BN	64.7 %	14.9 %	71.1 %	64.7 %	65.1 %
SVM	62.6 %	18.1 %	63.9 %	62.6 %	62.8 %
k-NN	81.4 %	9.8 %	81.4 %	81.4 %	81.4 %
DT	78.3 %	11.2 %	78.4 %	78.3 %	78.3 %
MLP	70.8 %	14.2 %	72.4 %	70.8 %	71 %
Logistic	61.3 %	18.8 %	62.5 %	61.3 %	61.5 %

0.614 for sticking. The k-NN algorithm exhibited a higher precision than other algorithms. In case of healthy adults, the average precision for all gait patterns was 0.974. The precision for all gait patterns was greater than 0.9. Moreover, the precision of all classifiers was greater than 95 %. Among them, the precision of the k-NN algorithm was higher than the other algorithms.

The k-NN algorithm exhibited a higher precision than the other classification algorithms with values of 99.5 % and 81.4 % for healthy adults and elderly people, respectively. The accuracy of the k-NN algorithm is significantly reduced when noise or unrelated features are present, and the feature size does not match the importance. The performance of the k-NN classifier is primarily determined by the choice of k as well as the distance metric applied [7]. Generally, larger values of k are more immune to noise, but the boundary between classes becomes unclear. We were able to improve the classification accuracy by setting the value of k to one. Moreover, we combined the two distinguishing features of hobbling of the left and right legs into one feature called "hobbling" in order to improve gait-pattern classification.

4 Conclusion and Discussion

The purpose of this study is to evaluate the difference in gait recognition between healthy adults and elderly people using the triaxial accelerometer sensor of a smartphone. Therefore, we extracted the features of three gait patterns and classified them using the triaxial accelerometer of a smartphone. The features were the mean, standard deviation, and horizontal and vertical components.

Nine elderly people over the age of 65 and 10 healthy adults were selected as subjects. In our previous study, "hobbling" was separately classified according to the left and right feet. However, it was difficult to separate these types because of the similar patterns and regions of accelerometer data. Thus, the two types of hobbling were combined into a single pattern.

For elderly people, the average accuracy about all gait patterns was 71.1 %. As a result, the precision of the k-NN algorithm was higher than the other classifiers at 0.814. In particular, the average accuracy for walking was higher than the average accuracy of sticking at 20 %. When elderly people used a cane the gait balance was irregular between each person. Thus, we think that this is the reason for the difference in the accuracy between walking and sticking. On the other hand, for healthy adults, the

average accuracy for the recognition of all gait patterns was 0.974. Among them, the k-NN algorithm was higher than the other classifier algorithms. The precision for healthy adults for all gait patterns was higher than that for elderly people by 24 %. In order to increase the precision for elderly people, we adopted various features. In the future, we plan to increase the precision by using gait-analysis-based data including the stride time, step count, step interval, peak-to-peak distance, etc. and increase the participate of subjects. Also, we plan to extract gait features to implement a fall prediction system. Gait classification using a smartphone can be used for disease and accident prevention for the future well-being of people. In particular, it can be used as basic research for a fall prediction system developed for elderly people. On the basis of the results of obtained in this study, it would be helpful to select an efficient classifier algorithm for gait analysis.

Acknowledgement. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1009) supervised by the IITP (Institute for Information & Communications Technology Promotion) and the Human Resource Training Program for Regional Innovation and Creativity through the Ministry of Education and National Research Foundation of Korea (NRF-2014H1C1A1066998).

References

1. Kannus, P., Parkkari, J., Niemi, S., Palvanen, M.: Fall-induced deaths among elderly people. Am. J. Public Health **95**(3), 422–424 (2005)
2. Jeon, M.Y., Jeong, H.C., Choe, M.A.: A study on the elderly patients hospitalized by the fracture from the fall. J. Korean Acad. Nurs. **31**(3), 443–453 (2001)
3. Chung, N.S., Choi, K.H.: Cause and prevention of falling in the elderly. Korean Acad. Univ. Trained Phys. Therapists **8**(3), 107–117 (2001)
4. Cheng, J., Chen, X., Shen, M.: A framework for daily activity monitoring and fall detection based on surface electromyography and accelerometer signals. IEEE J. Biomed. Health Inform. **17**(1), 38–45 (2012)
5. Ravi, N., Dandekar, N., Mysore, P., Littman, M.L.: Activity recognition from accelerometer (2005)
6. Dai, J., Bai, X., Yang, Z., Shen, Z., Xuan, D.: PerFalld: a pervasive fall detection system using mobile phones. In: 8th IEEE International Conference on Pervasive Computing and Communications Workshops, Mannheim, pp. 292–297 (2010)
7. Cover, T.M., Hart, P.E.: Nearest neighbor pattern classification. IEEE Trans. Inf. Theory **13** (1), 21–27 (1967)

Method of Detecting Malware Through Analysis of Opcodes Frequency with Machine Learning Technique

Sang-Uk Woo[1(✉)], Dong-Hee Kim[1], and Tai-Myoung Chung[2]

[1] Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon, Korea
{suwoo,kkim}@imtl.skku.ac.kr
[2] Department of Software, Sungkyunkwan University, Suwon, Korea
tmchung@skku.edu

Abstract. As the evolution of malware, vast damages are occurred in various industry fields. For this reason, research on malware detection has conducted actively. To improve the security of the network, SDN Quarantined Network (SQN) has been proposed. In this paper, we developed one of malware detection modules in first quarantine station in SQN by using the fact that benign and malicious files have different opcode frequency. And we applied machine learning technique as different way compare to conventional method. we verified that our module is valuable as one of detection modules and our final aim is to mount this module on the SQN system. Therefore, it would be possible more accurate inspection for new type of security attack with multiple detection modules.

Keywords: Malware detection · SDN quarantined network · Machine learning Operation code

1 Introduction

As the Information and Communication Technology (ICT) evolved, many devices connected to Internet have taken place throughout the industries. But still, modern computer system and communication infrastructures are vulnerable to various security attacks. Malicious software (Malware) is any software designed to interrupt normal operations, gather sensitive information, and generate immense damages in computer system. Many experts forecast that system will be damaged across a variety of industries. Even though there are many anti-malware solutions developed by private security companies, these solutions base their detection mechanisms on malicious behavior in the past and signature of malware. For this limitation, if an unknown type of security attack invades computer system, it is very difficult to counteract the incident without damages. To effectively cope with emerging sophisticated and complicated security attacks in advance, we need a new approach compared with conventional one. It is important to find the unique features of current malwares. There are several static analyses used for analyzing malwares (e.g. n-Gram, Byte Sequence, OPCODE, and

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_158

Portable Executable Header) [1]. Among them, we had noted the operation code (opcode) as one of the various features of malwares and malwares are confirmed to have a high frequency for rarely used opcodes. In computing, an opcode means a single instruction that specifies the operation to be performed. In this paper, we developed opcode module to detect executable files which have abnormal characteristic. Our module will be applied to First Quarantine Station in SQN which consists of several detection modules. With these multiple modules rather than only single opcode module, we can forecast that it is possible to accomplish more highly accurate inspection.

2 Related Work - SDN Quarantined Network

To defense thoroughly various security attacks, quarantine system was proposed with using SDN. It is called SDN Quarantined Network (SQN). The following describes explanation of static, dynamic analysis method and SQN architecture.

2.1 Static and Dynamic Analysis

The method of analyzing the malicious code can be roughly divided into three types: initial analysis, static analysis, and dynamic analysis. This section covers only static and dynamic analysis. First, static analysis is a method of analyzing a very specific operation by disassembling the malware file. Second, dynamic analysis is conducted by observing the behavior of the malware while it is actually running on a target system.

2.2 SQN Architecture

In this section, we explain SQN architecture. The SQN consists of five separate components (SQN Switch, Preprocessor, First Quarantine Station, Second Quarantine Station, and Clean Area Manager) as represented in Fig. 1. First of all, ingress traffic coming from external source arrives at the SQN Switch which conducts packet authentication procedure. According to processing result, the SQN Switch forwards authenticated packet to local network and also forwards non-authenticated packet to quarantine system which inspects packet in traffic coming from specific network source. The Preprocessor reassembles packets and forwards to the First Quarantine Station. The First Quarantine Station inspects whether incoming file (or data) has malicious characteristic by using static analysis. It is known that malicious data usually has characteristics like encryption, packing and unique operation code (opcode) frequency etc. There are several modules deployed in the First Quarantine Station to scan malicious packet in several phases because single detecting method is insufficient to inspect thoroughly. All modules are designed to reflect to features of malwares. The file gets into first module which checks data type. Then, the file is assigned to suitable path. For example, if the input data is executable file, it goes to PE header module and then arrives at opcode analysis module as Fig. 1. Each module gives score corresponding to inspection result. It is a kind of score system. If the score exceeds threshold value set in advance, the file will be dropped immediately in First Quarantine Station because it is

deemed dangerous file and some of indistinguishable files go to Second Quarantine System for the further inspection. Second Quarantine System conducts dynamic analysis and the rest of files which was decided to normal file are sent to Clean Area Manager with proper identification tag attached.

3 Implementation of Opcode Module

In this section, we introduce method of detecting malware through analysis of opcodes frequency with machine learning technique. In Daniel Bilar's paper, "Opcodes as predictor for malware", we can find that 14 opcodes most infrequently used is much better malware predictor than 14 opcodes most frequently used. In statistical analysis, infrequently used opcodes has more correlation to various malware classes (e.g. kernel-mode rootkit, trojan, virus, and worm etc.). For measuring correlation between opcode frequency and malware classes, value of Cramer's V is used. Cramer's V is a measure of association between two variables. Our opcode module disassembles sample files we collected into single opcodes separately, counts number of each opcodes and save the result as csv file. This csv file is a training data set. Therefore, by using training data set learning a large amount of sample data, unknown input file can be determined to whether it is malware or not. The following is a selected 28 opcodes [2] (Table 1).

We also had used the same kind of opcodes for detecting malwares and applied machine learning technique like decision tree to our opcode module. As decision tree is one of the popular machine learning techniques, it is useful classification method for categorical data and can analyze within a reasonable time on a typical computer

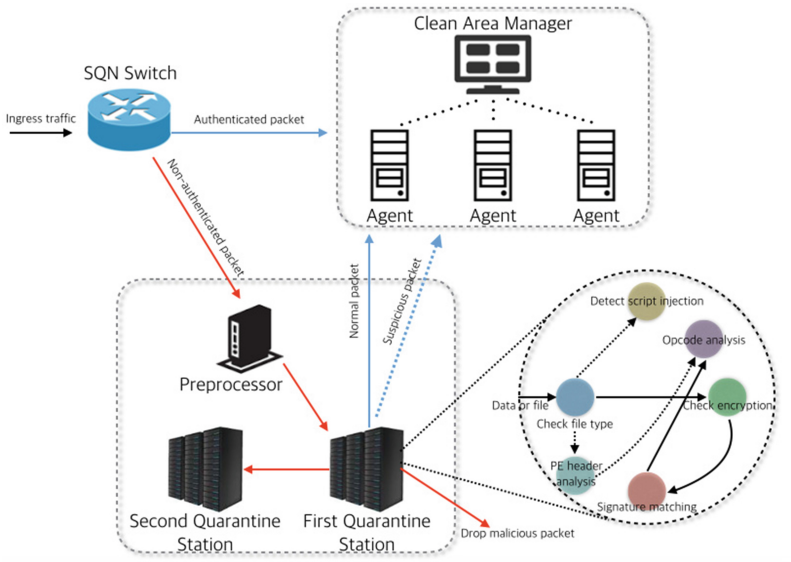


Fig. 1. SDN Quarantined Network (SQN) architecture

Table 1. The two kinds of opcodes used in implementation of opcode module

Most frequent 14 opcodes	mov	push	call	pop	cmp	jz	lea	test	jmp	add	jnz	retn	xor	and
Infrequent 14 opcodes	bt	fdivp	fild	fstcw	imul	into	nop	pushf	rdtsc	sbb	setb	setle	shld	std

environment. Classification and Regression Tree (CART) is one of the decision tree algorithms. Starting from entire data set, it can create two child nodes through splitting data set by using all predictor variables. Each root node indicates a single input variable and leaf nodes of the CART include an output variable which is used to make prediction [7]. We had to collect a vast amount of malicious and benign files to train the sample data and to find the optimal detection rate according to workload. We roughly collected more than 270,000 malicious files and 10,000 benign files. The source of benign files is win32 directory in Windows 7 and malicious files' one is VXheaven.org [8]. We configured test environment which has 11 training set. It consists of 10 training set by increasing the amount of learning data in 1000 unit. In last training set, we trained 100,000 malicious and 9,750 benign files to know changes in detection rate when malware's feature is more emphasized. We had to experiment continuously while changing the amount of learning data because the amount of the learning data can affect the detection rate and we put 1,000 malicious files selected randomly into opcode module for each 11 training sets. As a result of the several experiments, we confirm that opcode module can predict correctly malicious files and rate of false negative. In order to easily identify the results, we made a distinction between malicious and benign files by attaching tag like Fig. 2. The left digit 1 means that original input file is malicious file and the right digit 1 or 0 means result after the tests of the opcode module. Namely, the combination of 1 and 0 indicates false negative.

```
Trojan-Downloader.Win32.Agent.aioo,1,1
Trojan-GameThief.Win32.OnLineGames.hyi,1,1
Email-Flooder.Win32.Extreme,1,0
Trojan.Win32.Obfuscated.cfp,1,1
Backdoor.Win32.GrayBird.fb,1,1
```

Fig. 2. Example of results

In conclusion, we can classify whether the input file is benign or malicious as 94.3 %. The following Table 2 and graph in Fig. 3 describe the detail result according to the different amount of machine learning data. Even though the beginning result with 1,000 learning data shows that detection rate is low as 78.2 %, it gradually increases up to 94.3 % as learning data grow up to 109,750. Therefore, within the First Quarantine Station to perform multi-step inspection, our opcode module is valuable as one of modules in First Quarantine Station.

Table 2. Performance results for each 11 training sets

Amount of learning data (benign + malicious)	Detection rate (false negative/total)	Percentage (%)
1,000 (500 + 500)	218/1000	78.2
2,000 (1,000 + 1,000)	196/1000	80.4
3,000 (1,500 + 1,500)	167/1000	83.3
4,000 (2,000 + 2,000)	168/1000	83.2
5,000 (2,500 + 2,500)	174/1000	82.6
6,000 (3,000 + 3,000)	163/1000	83.7
7,000 (3,500 + 3,500)	148/1000	85.2
8,000 (4,000 + 4,000)	141/1000	85.9
9,000 (4,500 + 4,500)	132/1000	87.1
10,000 (5,000 + 5,000)	125/1000	87.5
109,750 (9750 + 100,000)	57/1000	94.3

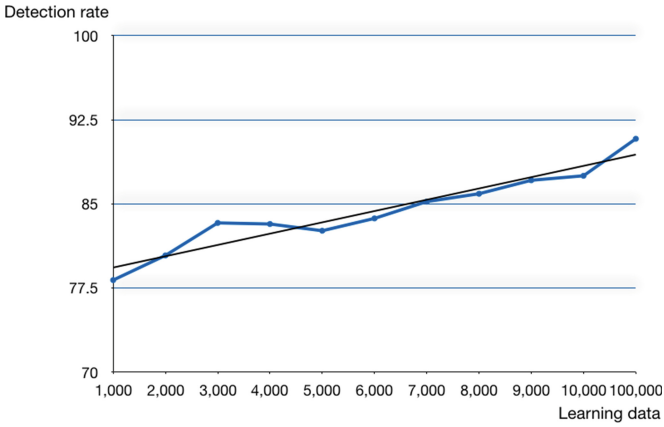


Fig. 3. Detection rate according as learning data grow up

4 Future Work

Until now, we only had conducted the tests on condition that the number of test sample is approximately 280,000 and the kind of opcodes is limited as twenty-eight. But, we suppose that there is an obvious correlation between detection rate and conditions mentioned above. In terms of some machine learning algorithms, the more learning data is used, the more we will get a good result. But, It's not always so. The algorithm like Support Vector Machine (SVM) shows low performance if when the excessive amount of learning data is trained. So, deciding the proper machine learning algorithm is as important as selecting optimal amount of learning data. In the future, we will study various machine learning algorithms, apply to our opcode module, and test with different kinds of opcodes for improving detection rate. In addition, we are looking

forward to developing another module to complete entire quarantine system as our ultimate goal.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2010-0020210).

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.R0113-15-0002, Automotive ICT based e-Call standardization and after-market device development).

References

1. Nath, H.V., Mehtre, B.M.: Static malware analysis using machine learning methods. In: Martínez Pérez, G., Thampi, S.M., Ko, R., Shu, L. (eds.) SNDS 2014. CCIS, vol. 420, pp. 440–450. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54525-2_39
2. Bilar, D.: Opcodes as predictor for malware. Int. J. Electron. Secur. Digital Forensics **1**(2), 156 (2007)
3. Santos, I., Brezo, F., Sanz, B., Laorden, C., Bringas, P.: Using opcode sequences in single-class learning to detect unknown malware. IET Inf. Secur. **5**(4), 220 (2011)
4. Shabtai, A., Moskovitch, R., Feher, C., Dolev, S., Elovici, Y.: Detecting unknown malicious code by applying classification techniques on OpCode patterns. Secur. Inf. **1**(1), 1 (2012)
5. Santos, I., Brezo, F., Ugarte-Pedrero, X., Bringas, P.: Opcode sequences as representation of executables for data-mining-based unknown malware detection. Inf. Sci. **231**, 64–82 (2013)
6. Santos, I., Nieves, J., Bringas, P.: Semi-supervised learning for unknown malware detection. In: Abraham, A., Corchado, J.M., González, S.R., De Paz Santana, J.F. (eds.) International Symposium on DCAI. AISC, vol. 91, pp. 415–422. Springer, Heidelberg (2011)
7. Decision Trees – scikit-learn 0.17.1 documentation. http://scikit-learn.org/stable/modules/tree.html
8. Vxheaven.org. Welcome to VX Heaven! (2016). http://vxheaven.org/

Study of Big Data Analysis Procedures

Joon Ho Park[1,2], Jin Ho Park[1,2(✉)], and Nam Young Lee[1,2]

[1] Department of Computer Science, Graduate University, Seoul, Korea
108.joonho.park@gmail.com, {j-park,nylee}@ssu.ac.kr
[2] School of Software, Soongsil University, Seoul, Korea

Abstract. In recent years companies and public institutions have long been accumulating and analyzing log data, and various types of data. It is building big data analytics platform for a variety of data analysis. Big data analysis platform may consist of distributed processing type quickly analysis large amounts of data. Recently, build a big data systems for companies and public institutions in the analysis of large amounts of data. Companies are providing a variety of services to users through data analysis. The public institutions are used to analysis a specific period of time and is useful in transportation, urban design, and commercial area analysis. In addition, national authorities to share that holds the data and the public and various analysis. However, it is a big data analytics projects that emerged as an issue in recent years this trend. Big Data Analytics project is to derive business results through data collection, data analysis and then build a big data platform. Currently, many companies have problems with the case of big data analysis projects using a methodology developed its own methodology and the CBD doing business.

In this paper defines big data Analysis Methodology, and suggests construct standards and construct procedures.

Keywords: Big data · Big data analysis · Big data methodology · Big data procedure

1 Introduction

In recent 2-3 years, companies and public institutions, in order to analyze large amounts of data, have done a big data system construction business. Companies and public institutions, after building big data analysis system, by data linkage of the data and other institutions, have a variety of big data analysis. Big Data analysis, when using analytical models initial design does not derive the result of the constant level, analysis and innumerable designs are repeated. Thus there is a need to develop a methodology that is suitable for big data analysis.

Big data analysis, in order to derive the result value of the objective from the vast amount of data, with the process of designing a big data analysis model, to verify and implementation.

Therefore, big data system construction business and big data analysis model project has recommended the Agile methodology.

In addition, the conventional ISO standards development methodology, program, has the advantage to implement database, user interface, the system operating environment.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_159

However, for use in big data analysis business there is a problem. Development methodology that domestic companies to use in the project of public institutions, are using a lot of methodology of the CBD methodology and information engineering. Software development business, development methodology of the CBD methodology and information engineering are suitable. However, big data analysis project, analysis, design, development, it is difficult to proceed to the phase of implementation.

Big Data analysis is to design a model that can generate information that can make decisions. Big Data analysis project, depending on the project period, analysis model of the design is performed, the detailed process, analysis, design models, decisions, is advanced from one week implemented within one month. A result of big data analysis, to apply the matters that have been determined through consultation with the customer in operation systems. There is a need for procedures to reflect the big data analysis model to the operating system.

In this paper, we present the analysis procedure that matches the characteristics of the big data analysis project.

2 Related Work

For the sake of the development methodology of the study of big data analysis, it presents the procedures and methods for big data analysis through the analysis of the development methodology being used in big data business, the advice of experts. Procedure Big Data analysis to determine the analyzed direction via the purpose of using analysis as a way to select the analysis target. In this paper, the definition and procedures for big data analysis, the entry of 9 created.

In this paper, to define the data that is to be analyzed as the basic data and the fusion data. Basic data is defined as data used for analysis of the data without processing. Fusion data defines the data reprocessing each basic data as a fusion data.

Execution of big data analysis, consists of three steps, for each procedure, is as follows. Step 1: ensuring the basic data and the fusion data. Step 2: Define a method of analyzing data. Step 3: result value derivation of data. Big Data analysis is statistical analysis depending on the type of analytical data, the density analysis, using the analysis techniques, such as linear analysis. In this chapter, to analyze the characteristics of big data analysis, presenting the analysis procedure.

2.1 Characteristics of Big Data Analytics

In this paper, in order to develop a procedure that is suitable for big data analysis, through the analysis of a plurality of big data business, it presents a new procedure and a model. Big data analysis requirements, unlike a typical software development requirements, in many cases there is no clear requirement. Thus, if not explicitly defines the requirements in the analysis phase has a high probability of fixed delay at the stage of analysis and design.

If big data analysis models designed requirement is in a state not clear, the result is changed based on the data type and analysis techniques. The type of data utilizing big

data analysis, the number, depending on the amount, is the value of the other results can be derived.

2.2 Big Data Analysis Procedures

Big data is analysis of the current situation rather than are focused on the future of the prediction and determination result accuracy. Therefore, the analysis processing speed of the data becomes timely, to derive a result. Big Data analysis is different from the general development projects repeatedly executing development with various designs. Big Data analysis was designed analytical model to derive the desired results from the enormous data, a process of validating implement. Big data analytics are not only to analysis business trends and trends for managers within the system, including the deployment of business representatives to support the decision-making using analytical models.

Procedure of big data analysis [Table 1], the definition of the problem, problem analysis, to progress through three stages of data definition.

Table 1. Big data analysis phase.

Division	Contents
Problem defined phase	Stylized in data analysis of the problem, it is necessary to set a hypothesis that can determine the success and failure
	Obtain data for solving the problem
	Another charge of business and the role selection participants
Problem analysis phase	Environmental analysis of the problem, business and analysis of relevant information infrastructure situation, set a target of analysis
Data definition	To understand the relationships between the collected from different data sources data, establish the concept of the business flow and data
	The correlation between the corresponding attribute for the purpose of the analysis is to select a high attribute
	The characteristics of the data attributes to be used as the input and output variables, the amount of data, to determine the analysis model taking into account the purpose of use
	Creating a data migration design document. (i.e.: Data Flow Diagram)

Big Data analysis, it is necessary to proceed after unambiguously determine the definition first problem, it is possible to reduce the repetition of the stage of design and implementation. In the data definition of the phase, data transfer design document may be selectively utilized in accordance with the characteristics of the big data analysis project.

3 Big Data Analysis Model Development Procedure and the Reference Design

To present the development deliverable guidelines for the development of big data analysis model. It needs to be created by each item presents a guide. The basic guidelines for the development of big data analysis model is defined in [Table 2].

Table 2. Excavation procedures and reference guide

Overview	Division	Contents
Procedures and criteria for excavation	Development procedure of application model	The main application requirements
	The definition of the purpose of the service	What to do analysis? (Clarity of purpose of the service)
	Information item confirm	What kind of data will you use the? (Use data definition, information held by institutions and business consultation)
	Apply technology review	Will to derive the deliverable of some of the analysis? (Selection of analytical methods, whether to build algorithm)
	Implementation scenarios review	What is the detailed service model building process and plans? (Detailed design work, detailed planning to establish whether)
	Construct effectiveness review	Is the operating model for building service desired effect is how much? (Ease of use, effectiveness, leverage Construct)

Finally, presented the items to be created through big data analysis guides [Table 3].

Table 3. Big data analysis guides

Division	Create content
Subject	Create a title challenge of analysis
Issues overview (required)	Need presentation if necessary to the problem of the analysis
Goal and differentiation proposal	The goal of the customer to the problem of analysis
Procedure of analytical methods	Create a procedure of analytical methods (To create the details of each step)
Utilization scheme	Utilization scheme presentation of the analysis results
Analysis target data	Data analysis (To be secured and ensure data is, data classification)
Visualization scheme	Visualization scheme presentation of the analysis results
Expected effect	Expected effect presented through the results of the analysis
Utilization of plan	Utilization scheme presentation of the derived results

4 Conclusion

In this paper, on paper of the limit, the evaluation method of the big data analysis, was not presented. In addition, the analysis procedure to be presented in this paper, the verification is applied to big data project in the future. It is derived through a validated process modification completion, to be presented in the following research paper.

So that it can be used in big data analysis project, you need to configure the template in consideration of the various cases. Procedures for big data analysis, some of the presents the development deliverable so that it can be used in the development methodology, whether it is possible to give a quantitative effect throughout the project execution, there is a need to develop a quantitative indicators.

Finally, use the big data analysis procedure presented in this paper, the delay of the analytical work that occurred in the existing Big data analysis business wants is improved.

References

1. Choi, E.H., Hwang, H.S., Kim, C.S.: Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces. Int. J. Inf. Commun. Eng. **9**(4), 358–362 (2011)
2. Proakis, J.G.: Digital Communications, 4th edn. McGraw-Hill, New York (1993)
3. Hennessy, J.L., Patterson, D.A.: Instruction-level parallelism and its exploitation. In: Computer Architecture: A Quantitative Approach, 4th edn., pp. 66–153. Morgan Kaufmann Pub., San Francisco (2007)
4. Hashmi, A., Berry, H., Temam, O., Lipasti, M.: Automatic abstraction and fault tolerance in cortical microachitectures. In: Proceeding of the 38th Annual International Symposium on Computer Architecture, New York, pp. 1–10 (2011)
5. Alavi, B.: Distance measurement error modeling for time-of-arrival based indoor geolocation. Ph.D. dissertation, Worcester Polytechnic Institute, Worcester, MA (2006)
6. Ben, Y.Z., John, D.K., Anthony: Tapestry: an infrastructure for fault-tolerant wide-area location and routing. University of California, Berkeley, CA, Technical Report CSD-01-1141 (2001)
7. Malardalen Real-Time Research Center, The worst-case execution time (WCET) analysis project. http://www.mrtc.mdh.se/projects/wcet/
8. Nowakowska, H., Jasinski, M., Debicki, P.S., Mizeraczyk, J.: Numerical analysis and optimization of power coupling efficiency in waveguide-based microwave plasma source. IEEE Trans. Plasma Sci. **39**(10), 1935–1942 (2011). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6003795

Author Biographies

Joon ho park received his bachelor's degree of Mobile Communication Engineering in Dongyang University, Gyeongbuk (2003). He is Master Degree of Information Science in Graduated Korea National Open University, Seoul (2015). He is studying Ph.D. software engineering in Graduated Soongsil University, Seoul. From 2016 until now, from the DAEBO Information Communication work in the Big data analysis, CF Algorithm Development Methodology Process, the main areas of interest software engineering, development methodologies, Big data analysis, etc.

Jin-Ho Park received his bachelor's degree of Software Engineering in Soongsil University, Seoul (1998). And master's degree (2001), doctor's degree of Computer Science in Soongsil University, Seoul (2011). Now he is a professor in the School of Software, Soongsil University, Seoul, Korea. His research interests focus on Software Engineering, SW Safety/QA/Testing, SW Convergence/Power, IoT, National Defense ISR, IT Service, IT Technical Commercialization and Start-up, etc.

Nam-Yong Lee received his bachelor's degree of Computer Science in Soongsil University, Seoul and master's degree of MIS in Graduated Korea University, Seoul. Doctor's degree of MIS in Mississippi State University, USA. Now he is a professor in the School of Software, Soongsil University, and Seoul, Korea. His research interests focus on System Engineering, Software Engineering, E-Commerce System, MIS, etc.

Design and Implementation of Authentication Information Synchronization System for Providing Stability and Mobility of Wireless Authentication

Yong-hwan Jung[1,2], Jang-won Choi[1], Hyung-ju Lee[1], Joon-Min Gil[2], and Haeng-gon Lee[1(✉)]

[1] Advanced KREONET Center, Korea Institute of Science and Technology Information, Daejeon, Korea
{paul7931, jwchoi, lhj273, hglee}@kisti.re.kr
[2] School of Information Techology Engineering, Catholic University of Daegu, Gyeongsan, Korea
jmgil@cu.ac.kr

Abstract. According to increasing the wireless network infrastructure and diffusion of mobile devices, the education environments equipped with mobile devices are gradually spreading in the field. The basic method to support stable wireless services in these education environments is to use wireless authentication technologies. The current education environments in Korea have been provided wireless authentication services with only unit of local areas. Accordingly, users cannot access the wireless network infrastructure for education in other areas outside local areas and thus the infrastructure is vulnerable to failures due to the lack of resource management and the absence of a backup authentication system for entire areas. In this paper, we suggest a Authentication Information Synchronization System (AISS) for stability and mobility.

Keywords: Wireless · Authentication · Synchronization · Stability · Mobility

1 Introduction

With the advanced technologies of wireless networks and mobile devices, recently, the distribution of mobile devices such as smartphones and smart pads has expanded. In an education field as well, the use of mobile device has become more common [1, 2]. One of the most fundamental technologies needed to provide stable education environment using a wireless network is the wireless network authentication technology under which only authorized persons are able to get access to the network. However, the current wireless network infrastructure in schools are poor in terms of stability and mobility due to the following reasons: absence of backup authentication, establishment of different wireless network infrastructure by region and inhibition on the authentication of the users from other regions. To solve this problem, we propose an authentication information synchronization system which can build a centralized backup

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_160

authentication system to provide wireless network environment for education, guaranteeing stability and mobility in local schools.

This paper is organized as follows: In Sect. 2, current wireless authentication environment which has been established in schools and its limitations are reviewed. In Sect. 3, the authentication information synchronization system is analyzed. In Sect. 4, a test on the proposed system and its results are stated. Section 5 provides the implications of the results and their utilization plan.

2 Related Works

2.1 Current Wireless Infrastructure for Education

At present, wireless network infrastructure has been built in schools across 17 regions. Depending on each region's budget and management staff, however, different wireless equipment (e.g., access point, access point controller, wireless authentication system, etc.) has been introduced and established. Even some of the regions where a wireless authentication system is established do not have a backup wireless authentication system. Therefore, if the current authentication system fails, there is no way to get access to the wireless network. Furthermore, there are regions in which the construction of wireless network infrastructure has been delayed. In addition, a low-price wireless router has even been installed without a separate authentication system. In this case, it is very vulnerable to a security breach. In other words, it can hardly prevent an authorized person from getting access to the internal system. As described above, even if there are the diverse forms of wireless network infrastructure, users should be able to get a stable and continuous wireless network for the high quality of services in education.

2.2 Constraints and Technical Considerations

The technical constraints needed to establish a backup wireless authentication system can be divided into three categories. First, it is needed to select an authentication method. There are many user authentication methods such as ID/PW, public key certificate and i-pin [3–6]. However, this paper considered a backup authentication system for the ID/password-based wireless authentication system. Second, there should be diverse supports on wireless network infrastructure. The hardware and software for the wireless authentication system differ by region so that a separate web application server (WAS) designed for a backup authentication system was considered in this paper. Third, there is a constraint in mobility. At present, wireless authentication provides wireless network environment against users within the related region only. A user from the other region is recognized as an unauthorized person so that he/she may not be able to get access to the wireless network. He/she may be authorized to get access to the system as a new user or under poor security. Because this kind of limitation in mobility is against the latest education paradigm, there should be a way to get access to the wireless network for education without time and space constraints. For this, this paper considered hierarchical wireless authentication based on a backup authentication system.

3 Design and Implementation of Authentication Information Synchronization System

3.1 Comparison Between the Proposed System and Conventional System

The functions of the authentication information synchronization system proposed in this paper can be divided into stability and mobility. Table 1 shows functional differences between the proposed and conventional systems.

Table 1. Comparison (Proposed system vs. Conventional system)

Category	Conventional system	Proposed system
Stability	Absence of the backup authentication system	Wireless authentication in a region in which no authentication system is available Each region's authentication database through security channels and regular polling-based synchronization Backup authentication if the system fails
Mobility	Wireless authentication within the region	Authentication of the users from other regions through hierarchical authentication Wireless authentication, using the unilaterally encrypted password

3.2 Design of Authentication Information Synchronization System for Backup Authentication

The proposed system provides wireless authentication services in the regions where a wireless authentication system is unavailable and offers backup authentication in all regions. For this, an authentication information synchronization system comprised of three modules (information collection module, database linkage module, file management module) has been designed and implemented. Then, user authentication information is synchronized through periodic file (CSV)-based polling synchronization process in the central control center and each region. Then, all communications between wireless authentication synchronization systems in the central and regional control centers are provided through safe and secure channels. Figure 1 and Table 2 show the detailed diagram of the authentication information synchronization system and its functions for each module.

3.3 Definition of Essential Synchronization Information for Backup Authentication

The wireless authentication system in the central control center established for the purpose of backup authentication should accept all users' authentication information. Because authentication information should be synchronized with the authentication

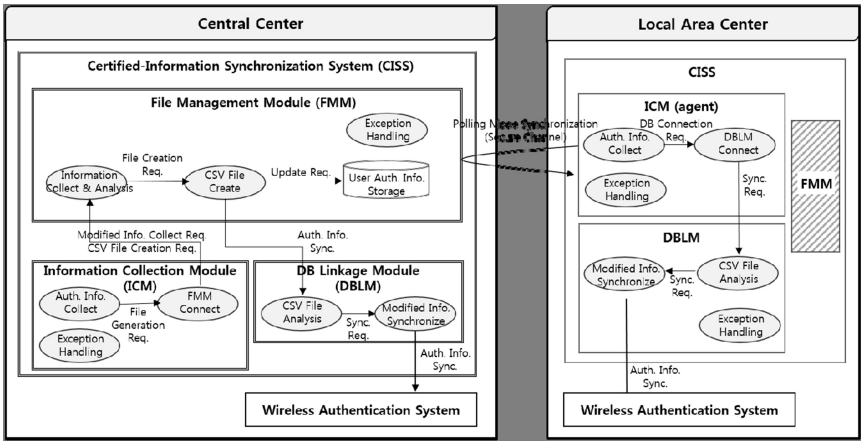


Fig. 1. Detailed diagram of authentication information synchronization system

Table 2. Detailed diagram of authentication information synchronization

Category	Functions
Information Collection Module (ICM)	Collects information on file management modules in other authentication information synchronization systems Sets the information collection cycle Classifies the collected information by region and information element Sends the collected information to the file management module Sends a synchronization order to other authentication information synchronization systems
DB Linkage Module (DBLM)	Links with the current wireless authentication database Loads the collection information on the wireless authentication database Sets the read/write authority needed to get access to the database Saves and manages the history of synchronization
File Management Module (FMM)	Creates the collected information in CSV file format Manages the authentication information adjusted by region in each file Plays as a CSV temporary storage

system in each region regularly, the essential information for synchronization needs to be minimized. These essential data must be applied to the authentication system database in all regions. Table 3 shows the essential synchronization items that must be available for backup authentication.

Table 3. Essential synchronization database composition items

Category	Essential synchronization items
User information	ID, password, user type, school name, name, institution code, grade, class, ID
Wireless equipment information	Institution name, equipment type, IP, encryption key, MAC

4 Experimental Test and Results

4.1 Experimental Test and Results of Authentication Information Synchronization

In an authentication information synchronization test, the user authentication information registered in the wireless authentication system in the central control center is synchronized in the wireless authentication system in two regional control centers. Therefore, the test was performed under a scenario which determines the accessibility to the regional wireless network environment based on the authentication information. The confirmation process and results by the test stage are as follows:

① Registration of new user authentication information from two regional centers in the central authentication system.
② Synchronization of authentication information through the authentication information synchronization system (Fig. 2).
③ Confirmation of user authentication information synchronization in the authentication system from two regional centers (Figs. 3 and 4).

Fig. 2. Registration of new user authentication information in the central authentication system

```
MariaDB [syncagdb]> select * from s_sync_user_info;
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+----------------+----------------+
| aaa_server_seq | user_name | user_password                                | office_edu | local_office_edu
| is_change_flag | register_dt    | update_dt      |
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+----------------+----------------+
|              0 | test10    | 7Cc4/rK7sLx4PrRmeQM5FBY3K6bti43dvrvbN+UQJHM= | dje.go.kr  | dje.go.kr
|                | 20140418011949 | 20140423035408 |
|              0 | test9     | tEUQNNO2WQBgzpSEoouI3TMqgKIq6OOcnFy3NXqybJ8= | dje.go.kr  | dje.go.kr
|                | 20140418011949 | 20140423035408 |
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+----------------+----------------+
2 rows in set (0.00 sec)
```

Fig. 3. Confirmation of authentication information synchronization in the authentication system in 'A' center

```
MariaDB [syncagdb]> select * from s_sync_user_info;
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+
| aaa_server_seq | user_name | user_password                                | office_edu | local_office_edu |
| update_dt      |
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+
|              0 | test21    | f/3Wz4Ym+nJ8Xw328ML/qEJssBddvkYWveXfAXt3Yv0= | gne.go.kr  | gne.go.kr        |
| 20140422191044 |
|              0 | test22    | j8XabiWfbkawksNMuc7V+Z6Ow+Bxd4rykE240bzwiCs= | gne.go.kr  | gne.go.kr        |
| 20140422191044 |
+---------------+----------+-------------------------------------------+-----------+-----------------
+---------------+
2 rows in set (0.00 sec)
```

Fig. 4. Confirmation of the synchronized authentication information of the authentication system in 'B' center

④ Confirmation of wireless user authentication through a mobile device (Fig. 5).

Fig. 5. Confirmation of wireless user authentication through mobile device

4.2 Experimental Test and Results of Backup Authentication

The backup authentication test was conducted under a scenario in which a user tries to get access to the wireless network after the wireless authentication system in the regional control center is shut down. The confirmation process and results by the test stage are as follows:

① Discontinuance of the wireless authentication system in the regional center (deemed as 'fault') (Fig. 6).
② Check on the authentication in the central wireless authentication system in the event of an attempt to get access to the regional wireless network (Fig. 7).

Fig. 6. Discontinuance of regional wireless authentication system

Fig. 7. Confirmation of backup authentication in the central wireless authentication processing system

5 Conclusion

In this paper, we proposed an authentication information synchronization system with a goal of establishing wireless network authentication environment which guarantees stability and mobility optimized for school environment. The proposed system is able to immediately respond to any failures of wireless authentication system in the regional control centers, making stable and continuous wireless network authentication services possible. In addition, it is able to build a hierarchical wireless authentication system which ensures user mobility through simple proxy adjustment in the regional wireless authentication system. As a result, users would be able to get education easily anytime and anywhere without time and space constraints, using the wireless network for education.

References

1. NIA, School of advanced wired and wireless infrastructure plan for smart education (2012)
2. KISTI, Guidelines for Wireless Network Authentication System and associated standard (2013)
3. RFC 5281, Extensible Authentication Protocol Tuneled Transport Layer Security Authenticated Protocol Version 0 (EAP-TTLSv0), IETF (2008)

4. PAP. http://en.wikipedia.org/wiki/Password_authentication_protocol
5. RFC 2284, PPP Extensible Authentication Protocol (EAP), IETF (1998)
6. RFC 7268, RADIUS Attributes for IEEE 802 Networks, IETF (2014)

Study on the Generic Architecture Design of IoT Platforms

Mi Kim[1(✉)], Nam Yong Lee[2], and Jin Ho Park[2]

[1] Department of Computer Science, Graduate School, Seoul, Korea
pytwoori@gmail.com
[2] School of Software, SoongSil University, Seoul, Korea
{nylee, j.park}@ssu.ac.kr

Abstract. A variety of application platform resulted in a variety of requirements that IoT systems should comply with. Due to the heterogeneity of the environments, the requirements varied significantly, and demanding more or less complex systems with varied performance expectations. This situation affected the architecture design and resulted in a range of IoT architectures with not only configuration setting of IoT devices and resources, but also varied environments of collaboration each devices. A number of things connecting it proposes a generic architecture design platform to useful the common characteristics which internet of things to control the flow of data.

Keywords: IoT applications · Internet of things · Architecture design · Platform · Framework

1 Introduction

The Generic IoT platforms can be used with good results in participation with device of an IoT environmental applications. The IoT platform consists of the Service application and environment monitoring service capabilities and IoT applications functions. The goal of this paper is Generic architecture design for IoT platforms in environments collaboration devices and user runtime environments, trying to guarantee the consistency as well as effectively energy consumption. It can be integrated heterogeneous and strict requirements, such as limited resource time and battery. The IoT platform is proposed so as to build an IoT-based Configurable resource time and battery consumption Service Platform for design. A number of things connecting it proposes a generic architecture design platform to useful the common characteristics which internet of things to control the flow of data. The rest of this paper is organized as follows and discusses the related work, Sects. 2, 3 is presents a design for IoT Platform and evaluation and Sect. 4 gives a conclusion. Most IoT management platforms, such as Philips hue [1] and IoT.est [2], focus on home automations or sensor networks. These platforms have shown excellent performance as they claimed. Philips hue is a personal wireless lighting system which can only be controlled by a smart device (e.g., smartphone). to more than 4.5 billion USD by the end of 2011 thanks to the introduction of open APIs [3]. The solutions for IoT implementation focus on uniform naming [4] and addressing, and common protocols for ubiquitous smart objects.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_161

It concerns objects reachable both directly by IP network or connected using dedicated protocols specific for given area of use. This approach makes feasible the creation of common platforms which aim to provide tools for management, data analysis and adequate reaction, and makes use of numerous IoT middleware solutions providing interoperability. center on exposing virtual representations of physical objects in order to develop a common communication platform. In this way, SENSEI project [5] introduces the concept of resource which corresponds to a physical entity in the real world. Summarized, while there are several papers available describing what an existing IoT platform should offer from a special structure for platform, hardly any domain is available on the common capabilities of existing IoT platforms. However, relevant research builds upon relevant IoT application problems.

2 Design of Generic IoT Platforms

In this section, a design for IoT platform is consists of proposed to framework for generic architecture design model in section.

2.1 Design of Generic IoT Platforms

The design for IoT platform is consists of IoT generic service Interface, it can be connected between IoT service and IoT platform as shown in Fig. 1.

Generic driver interface also connected with IoT platform which composite API components are meta model, device manager, configuration manager and security Manager.

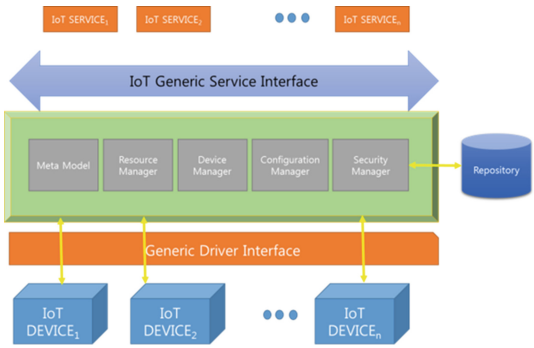


Fig. 1. Architecture for IoT platform

2.2 IoT Generic Service Interface Algorithm

In this section, I suggest the IoT generic service interface for heterogeneous interface. It can be invoke a method through one interface from heterogeneous Services. To satisfy this method I applied generic interface parameter on platform.

This algorithm is to represent generic interface for heterogeneous IoT Service. The Service has various application for drone, wearable watch and smart car service etc. Next table [Table 1] shows a generic service interface algorithm on the IoT platform.

Table 1. Generic service interface algorithm

```
1.  Step1 Internal  communication Interface//
2.  Input data  : UserID, Configuration data
3.  Output data : service data
4.  Target : IoT Service
5.  ----------------------------------------------------------------
6.  Method : IoT_Generic_Service// TRANSFER
7.  // step 1.  Set Internal Interface
8.  Try {
9.  Request data type assign to common global type // request data type is
    variability
10. Set Req1 assign to userid_req(IoT Service)  // set of userid request
11. Set Req2 assign  to sensory_req(IoT Service) // set of sensory_request
    data2
12. Set Trans assign to req.trans // set a transfer method in request
13. ----------------------------------------------------------------------
14. // step 2 External communication Interface // External communication for
    IoT Service
15. Set Trans assign to  Resensor(data)
16. For read  Trans
17.     IF(Trans =! null) // trans data is not null
18.         Trans = GenericRequest // request
19.         Count = Count +1;
20. Mapping Response = Resensor // mapping Response data and request data
21. Transfer  Response // data transmission
22. }
```

2.3 Generic Driver Interface

The Device has various sensors and actuators. Next table [Table 2] is a generic driver interface algorithm on the IoT platforms.

2.4 Configuration MANAGER

The collaboration manager flow assumes that IoT device information such as device name, id, or description already stored in the repository of IoT platforms. First, as Initialize IoT device connects with IoT collaboration device. After that, Check the Available Devices and Check Liveness executes to check the connected device is available in current.

If it is not available device to send failure message and is alive it should assign to collaboration device such as finding another available device (Figs. 2 and 3).

Table 2. Generic driver interface algorithm

```
1 public classGeneric_driverInterface
2 ----------------------------------------------------------------------
3 public intconnectDevice() {
4 Try {
5 Initializing   Generic Driver Interface
6 Search Generic_driver_connection
7 Set Generic_driver_connection equal to discovery driver1
8 Mapping parameter driver1 interface param
9 Print driver1
10 For each driver   assign to driver interface
11        If driver connection to device
12             then success device connection
13             Return ;
14 Search next device driver
15          For each next driver
16               If next device is discovery
17                    Then next driver   assign to device
18           Connection close device
19   Return;
```

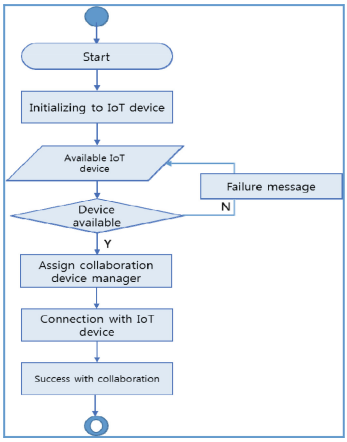


Fig. 2. Collaboration manager

2.5 Security Manager

Secure execution environments is it refers to managed code, runtime environment design to protected IoT platforms. Secure for network provides a network connection or service access only if the device is authorized. Security manager includes authenticating communicating peers, ensuring confidentiality and integrity of communicated data preventing repudiation of communication of transaction and protecting the identity of communicated entity as sensory data, resources. Secure for repository involves confidentiality and integrity of sensitive information stored in the system.

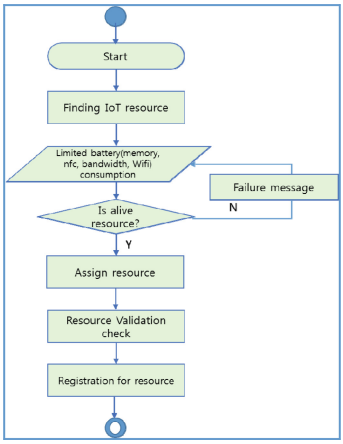


Fig. 3. Resource manager process

3 A Case Study and Evaluation

3.1 A Case Study

To show the applicability of the proposed design of IoT platforms. We conduct a case study of applying the experiments and scenario an IoT Platforms.

We demonstrate the usefulness of we suggested generic architecture for IoT Platform and the above architecture using a case study.

We have suggested a scenario where mobile application can be controlled according to the IoT monitoring service. It is assumed that the discover the special device driver to the end user contains discovering a light sensor and a temperature sensor. These sensors are connected to an IoT Platform. The IoT device first queries the platform to discover a driver set of sensors. In this case, the discovery reveals that the IoT monitoring service has a light sensor and a temperature sensor and both of these sensors are linked to IoT platform. If the user selects temperature sensor and weather domain, then the IoT monitoring service the IoT repository to obtain a list of the available device driver. The list consists of four such scenarios: (i) Weather, Luminosity and Emotion, (ii) Weather and wearable watch, (iii) Weather and Activities and (iv) Weather, delivery service and Safety Device. If the user chooses the last option, based on the temperature sensor measurement, the IoT monitoring Service can deduce whether the authenticity device is real.

According to the deduction, the user receive the service of exact whether temperature. The generic interface can provide a common set for functionality.

We suppose scenario of generic IoT Platform can be easily discover the real device driver specially.

3.2 Evaluation of IoT Platform

We have evaluated the performance of the generic IoT platform in terms of CPU load and power consumption. These two metrics are very important since the application is

running on IoT Monitoring devices with limited battery life. The application has been tested on three other platform. Table 3 lists the CPU load results as measured from DDMS tool of Eclipse ADT. The loads are measured during the main four operations phases of the application.

Table 3. CPU usage of the IoT platform

Division	Generic IoT	EVRYTHNG	THINGWORX
Discovery	3 %	–	–
Actuation	1 %	3 %	5 %

The IoT service operations and the CPU computations take minimal power. The dissipation at the display can be reduced by lowering the brightness value of the smart device.

Table 4 lists the power consumption results as measured from power tool of Eclipse ADT.

Table 4. Power consumption of Platform (mW)

Division	Power consumption (mW)	Division	Power consumption (mW)
	Generic IoT	EVRYTHNG	THINGWORX
Sensor data	277	Sensor data	277
WIFI	205	321	255

4 Conclusion

In this paper, this variety of application platform resulted in a variety of requirements that IoT systems should comply with. Due to the heterogeneity of the environments, the requirements varied significantly, and demanding more or less complex systems with varied performance expectations. This situation affected the architecture design and resulted in a range of IoT architectures with not only configuration setting of IoT devices and resources, but also varied environments of collaboration each devices. A number of things connecting it proposes a generic architecture design platform to useful the common characteristics which internet of things to control the flow of data. In the future work, a suggested generic architecture design can apply with case study results of CPU load and Power Consumption of IoT platforms. It will be comparison of other platforms a Generic IoT platform it can be measure relatively accurately.

References

1. Philips hue. “Meet hue” (2014). http://www.developers.meethue.com/
2. De, S., Carrez, F., Reetz, E., Tönjes, R., Wang, W.: Test-enabled architecture for IoT service creation and provisioning. In: Galis, A., Gavras, A. (eds.) FIA 2013. LNCS, vol. 7858, pp. 233–245. Springer, Heidelberg (2013). doi:10.1007/978-3-642-38082-2_20

3. Voxeo Labs Tropo whitepaper: Make the Shift From Telco Power to Telco Powered with the Tropo API 2013
4. MongayBatalla, J., Krawiec, P., Gajewski, M., Sienkiewicz, K.: ID layer for internet of things based on name-oriented networking. J. Telecommun. Inf. Technol. **2013**(2), 40–48 (2013)
5. EU FP7 SENSEI Project Consortium: Final SENSEI Architecture Framework. SENSEI Project Deliverable D3.6 (Checklist of Items to be Sent to Volume Editors)

A Study on Digitalization of Seafarer's Book Republic of Korea for e-Navigation: Focusing on Wireless Network

Jun-Ho Huh(✉)

Research Professor (Computer Engineering Ph.D), Department of Architectural Engineering, Dankook University at Jukjeon, Yongin, Republic of Korea
72networks@pukyong.ac.kr

Abstract. Most of nations including the Republic of Korea require that the captain or the chief engineer of the ship who is in charge of safe navigation must carry a Seafarer's certificate. However, different from the merchant seamen, a considerable number of Korean fishermen do not follow such a rule as the procedure to obtain the certificate is too inconvenient so that the certification system is in a way creating offenders unintentionally. Thus, this study focuses on digitalization of Seafarer's certificate to adapt to forthcoming e-Navigation era. As a result/contribution, it became clear that the current BLE technology (2016) was most efficient in achieving this goal among various wireless network-oriented technologies. The author intends to introduce a prototype in the future study after applying for a patent.

Keywords: Digitalization · Bluetooth · BLE · Seafarer's Book e-Navigation

1 Introduction

This study has focused on the IDs cards which have been used widely to authenticate the members of certain groups or organizations. Many firms around the world are now monitoring staffs's daily routine and reinforcing the security by digitalizing their IDs.

The captain or the chief engineer responsible for the safety of crews and the vessel must carry a Seafarer's certificate which is often used as a passport on the ocean freighters [1–5]. The fishing boats working at the littoral seas do not require the certificate and yet, the captains and chief engineers are required to gain an official approval from the Regional Office of Maritime Affairs and Fisheries every time they change the vessel [2–7]. After the introduction of the Seafarers' Identity Documents Convention (ILO) into the Seamen Law (ROK), existing Seafarer's Certificate is now being used to document Seafarer's on-board working experiences, qualifications, employment contracts, and etc. However, the certificate cannot be admitted as a passport and the recent immigration control law in the ROK requires all the foreign seamen to carry the Seafarer's Identity Documents when they enter or leave Korean ports. This law started to take an effect on the 1st day of June 2017 so that the seamen had to carry their passports until they each received a new Seafarers' Identity Documents. Figure 1 shows design of current (2016) Korean Seafarer's Book.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_162

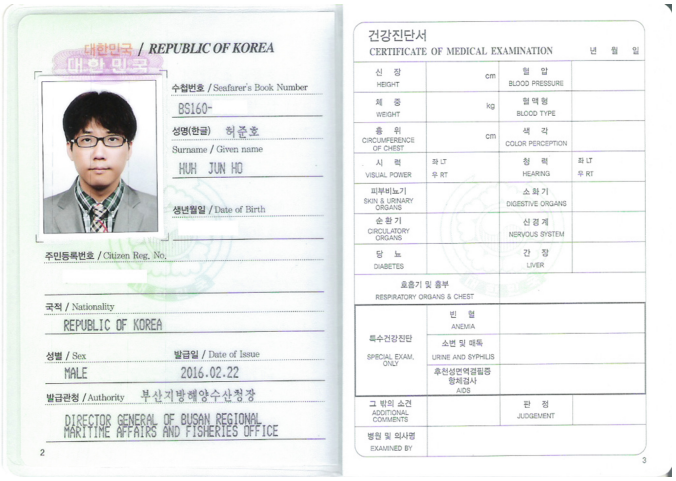


Fig. 1. Design of current (2016) Korean Seafarer's Book

There were many reasons for not being able to digitize the Seafarer's Certificate but one of the major reasons was that since the RFID chips used for the existing electronic passports were vulnerable to the harsh on-board environments (e.g., extreme temperature and humidity), they couldn't be used for the Seafarer's Book.

2 Related Research

2.1 Electronic Passport (ePassport)

The electronic passport, or ePassport, refer to the machine-readable passports embedded with an antenna and a chip containing carrier's personal and biometric information. The original purpose of electronic passport is to optimize the security of passports through prevention of forging, falsifying or appropriating to ultimately increase the convenience of Korean nationals traveling abroad.

The chip used for the ePassport contains the same information stored in the old-style passport but some security-related technologies are added to it. Thus, it is quite difficult to forge the information printed on the passport pages and contained in the chip simultaneously. Even if it is possible to do so, manipulations will be automatically discovered when entering/leaving the country. The ROK government is issuing diplomatic passports and general passports in the form of ePassport since Mar. 31st 2008 and Aug. 25th 2008, respectively [8–12].

Few problems involved in manufacturing ePassports are that the passport's cover will be thicker than the old types because of embedded antenna and chip and these two components can be damaged when the cover is bent by force, or stapled. One of the current security issues is related to the hacking of scanned passports. Many studies to protect scanned passport information have been introduced but it is unwise to be complacent yet. It is true that there have been some concerns about information leaking

through remote access as the non-contact chips are embedded in the ePassports. Accordingly, there also were many discussions and reviews on this problem by the experts prior to global introduction of the ePassport system. In order to prevent skimming, a technique that illegally steals personal information without being recognized by the passport holder, the security technology called BAC has been applied to Korean passports following the regulation of ICAO. This way, skimming becomes impossible when the passport is folded.

Further, the strong security measures such as the EAC and CA technologies are being applied to Korean passports to protect them from eaves dropping so that their security level is very high. Aside from these technologies, the ROK government is taking various measures (e.g., centralized passport issuing system, single closed passport administration network, and etc.) at an administrative level. The most difficult passports to forge are the Nicaraguan passports which have 89 security-related indicators. The wireless hacking of ePassports still remains as a global issue and the researches to replace RFID technology for passport use are ongoing.

2.2 Seafarer's Book

From 2004 to 2005, a plan to issue Seafarer's Book embedded with crew's biometric information was established but aborted because of several unsolved problems. The ministry of Maritime Affairs and Fisheries (MOF) explained in 2004 that this plan was based on the decision to adopt the Seafarers' Identity Documents Convention which obligated the seamen to carry a biometric Seafarer's certificate while the nations that issue such certificates must computerize the relevant information [12–16]. Following the 911 terror incident and the US government's decision that only the passports embedded with biometric information would be accepted starting from Oct. 26th 2005, the ILO also advised that the Seafarer's certificates that function as a passport should be changed accordingly. However, voice information was excluded in the new certificates and the major reason for delay in digitalization was that the RFID chips were vulnerable to rather extreme temperatures and humidity.

3 Digitalization of Seafarer's Book ROK for e-Navigation

There are more parameters in the Seafarer's certificate than ePassport Fig. 2. For example, the diagnostic information such as anemia, syphilis, AIDS and urine test results are included in Section (a) together with the hospital and doctor's names. And the Section (b) contains the details of the employment contract while Section (c) specifies the records of contract renewal(s) or term changes. The Section (d) includes the license type(s), validity(s), date of acquisition and other particulars while the Section (e) indicates holder's signature and address along with official requirements or instructions issued by the authorities. Any additional license(s) the holder has should be listed here. Author's previous study 'The Medicine(s) Intake Notification System' [1, 2] is used for the items included in the Section (a).

Fig. 2. Parameters in the Seafarer's Book than ePassport

The higher security and storage levels are required for this new type of certificate where many valuable personal information are being kept. These certificates are usually kept in the on-board safe by the captain but they sometimes get lost or stolen. The several experiments to distinguish damaged certificates due to a long-term use have been carried out in this study as well.

As described in Fig. 3, this study intends to track the position of a certain Seafarer's Certificate by triangulation based on the signals transmitted from a multiple number of beacons which act as the GPS satellites within a ship. This system employs existing principle of GPS operation and the Bluetooth Low Energy (BLE) technology. As mentioned earlier, the certificates are often kept in the safe but cannot be guaranteed

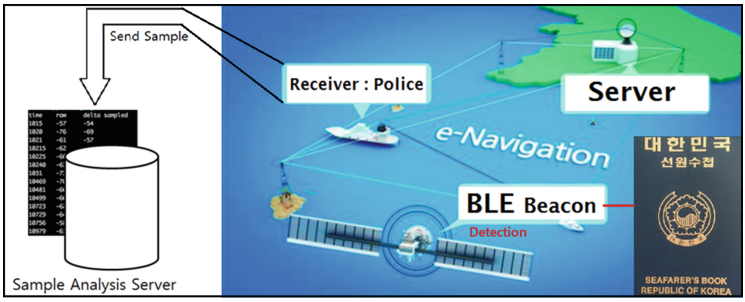


Fig. 3. A whole system design

from theft. Author's patent application will include a technology to recharge the BLE batteries within the safe.

Since the RSSI and the unique identifier of a BLE equipment can be identified, the position of a ceratin beacon can be determined using the latter and its distance between a ceratin beacon can be estimated with the former. Therefore, the position of a certificate holder can be deduced by triangulating the information transmitted from equipments scattered onboard. By using the Bluetooth 5 technology to check the certificates from outside of the ship, it will be possible for the cost guard or the investigative authorities to determine how many are working or boarding on the ship. Also, as the Bluetooth function is mounted on almost all the smartphones, they can avoid additional costs of purchasing new measuring equipments.

Meanwhile, the BLE-exclusive equipments can last about 1.8 to 28.7 months with a CR2045 battery because of their low-energy consumption rate. Figure 4 shows the BLE chipsets and their battery lives depending on the settings. No additional wiring works are required to install long-term-use beacons. The battery-recharging system introduced in this study will be quite useful for those Seafarer's Certificates with the validities of more than 5 years.

Chipset	Advertising Interval	Est. Battery Life CR2032	Est. Battery Life CR2045	Est. Battery Life CR2477
Gimbal	100ms	n/a	n/a	n/a
Gimbal	645ms	1 month	2.5 months	4.1 months
Gimbal	900ms	n/a	n/a	n/a
Nordic Semiconductors	100ms	1.2 months	3.1 months	5.1 months
Nordic Semiconductors	645ms	7.0 months	18.19 month	29.3 months
Nordic Semiconductors	900ms	11.1 months	28.7 months	46.29 months
Bluegiga	100ms	0.9 months	2.4 months	3.8 months
Bluegiga	645ms	5.9 months	15.4 months	24.8 months
Bluegiga	900ms	9.3 months	23.9 months	38.5 months
Texas Instruments	100ms	0.7 months	1.8 months	2.9 months
Texas Instruments	645ms	4.1 months	10.6 months	17.1 months
Texas Instruments	900ms	5.6 months	14.4 months	23.1 months

Aislelabs

Fig. 4. BLE and battery lives depending on settings

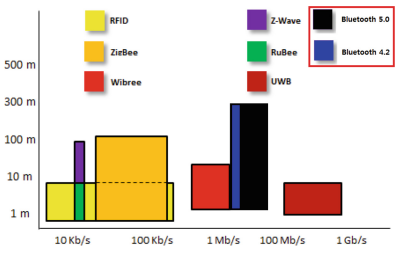


Fig. 5. Comparing Data Rate vs Data Range

Figure 5 shows the Data Rate vs. Data Range of each technology described in the preceding study [3]. Since the Bluetooth 5.0 and 4.2 indicated in the red box have good ranges and humidity superior humidity-resistancy, they are considered to be adequate in ship's environment.

On June 17th 2016. The Bluetooth Special Interest Group (Bluetooth SIG) announced the Bluetooth 5 technology. This technology has an improved transmission distance and better speed compared to other prior versions. Targeting the drone, wearable, smart bulb, home automation and IoT equipments markets, this newest Bluetooth technology in two years since the introduction of Bluetooth 4.1 in Dec. 2013 has increased its transmission distance four times longer and data transfer rate more than twice. It is expected that Bluetooth 5 will activate the service markets where information are provided based on the navigation or positioning technologies. According to the Bluetooth SIG, the products embedded with Bluetooth 5 will be launched in the first half of 2017. The author is proposing the Seafarer's Certificate

embedded with the BLE device prior to the one that will be using the Bluetooth 5 technology in the future.

4 Conclusion and Future Work

The researches on the digitalization of Seafarer's Certificates have been carried out in this paper to adapt to the future e-Navigation era. In search of the most adaptable sensors for the certificates, the author found that the Bluetooth 4.2 and BLE technologies were most efficient considering many factors.

The possibility of using the certificate as an e-Navigation system terminal has been pursued also pursued. Introducing a new terminal with which the batteries can be recharged within the cabin or the safe will be included in author's future work along with the prototype, after applying or registering for a patent. The processing and memory overheads are the major additional complications that could follow in our proposed protocol. As a future study, it may be possible to implement congestion control mechanism. The protection for the transmitted data was not taken into consideration in our protocol so that the discussion concerning the protection methodology can also be one of the future studies.

References

1. Huh, J.H., Kim, N., Seo, K.: Design and implementation of mobile medication-hour notification system with push service function. Int. J. Appl. Eng. Res. **11**(2), 1225–1231 (2016)
2. Huh, J.H., Kim, N., Seo, K.: Design and implementation of mobile push service-based mobile medication-hour notification system. ASTL, SERSC **117**, 92–96 (2015)
3. Zareei, M., Zarei, A., Budiarto, R., Omar, M.A.: Comparative study of short range wireless sensor network on high density networks. In: 17th Asia-Pacific Conference on Communications (APCC), pp. 247–252. IEEE (2011)
4. Ault, A., Coyle, E., Zhong, X.: K-nearest-neighbor analysis of received signal strength distance estimation across environments. In: Proceedings of the First Workshop on Wireless Network Measurements, pp. 1–6 (2005)
5. Jung, J., Kang, D., Bae, C.: Distance estimation of smart device using bluetooth. In: 8th International Conference on Systems and Networks Communications, pp. 13–18 (2013)
6. Dong, Q., Dargie, W.: Evaluation of the reliability of RSSI for indoor localization. In: IEEE International Conference, Clermont Ferrand, pp. 1–6 (2012)
7. Botta, M., Simek, M.: Adaptive distance estimation based on RSSI in 802.15 4 network. Radio Eng. **22**(4), 1162–1168 (2013)
8. Kang, J., Kim, D., Kim, Y.: RSS self-calibration protocol for WSN localization. In: IEEE 2nd International Symposium on Wireless Pervasive Computing, pp. 181–184 (2007)
9. Kumareson. P., Rajasekar. R., Prakasam. P.: Accurate location identification using bluetooth in android smartphone. Int. Daily J., India, 81–86 (2015)
10. Huh, J.H., Bu, Y., Seo, K.: Bluetooth-Tracing RSSI sampling method as basic technology of indoor localization for smart homes. Int. J. Smart Home, SERSC, Australia **10**(10), 1–14 (2016)

11. Huh, J.H., Seo, K.: RUDP design and implementation using OPNET simulation. In: Park, J. J.(J.H.), Stojmenovic, I., Jeong, H.Y., Yi, G. (eds.) CUTE 2014. LNEE, vol. 330, pp. 913–919. Springer, Heidelberg (2015)
12. Huh, J.H., Koh, T., Seo, K.: A shipboard secret ballot system for the ICT-isolated ocean crews. In: Park, D.-S., Chao, H.-C., Jeong, Y.-S., Park, J.J.(J.H.) (eds.) CUTE 2015. LNEE, vol. 373, pp. 549–557. Springer, Singapore (2015)
13. Eagle, N., Pentland, A.: Social serendipity: mobilizing social software. IEEE Pervasive Comput. **4**(2), 28–34 (2005)
14. Zhu J, Zeng K, Kim, K.H., Mohapatra P, Improving crowd-sourced wi-fi localization systems using bluetooth beacons. In: 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pp. 290–298 (2012)
15. Chang, Y.J., Chu, Y.Y., Chen, C.N., Wang, T.Y.: Mobile computing for indoor wayfinding based on Bluetooth sensors for individuals with cognitive impairments. In: 2008 IEEE 3rd International Symposium on Wireless Pervasive Computing, pp. 623–627 (2008)
16. Huh, J.H., Otgonchimeg, S., Seo, K.: Advanced metering infrastructure design and test bed experiment using intelligent agents: focusing on the PLC network base technology for Smart Grid system. J. Supercomput. **72**(5), 1862–1877 (2016). Springer, USA

A PMIPv6-Based Auxiliary Mobility Management Considering Traffic Locality

Ki-Sik Kong(✉)

Department of Multimedia, Namseoul University, Cheonan, Republic of Korea
kskong@nsu.ac.kr

Abstract. Proxy mobile IPv6 (PMIPv6), which is a centralized mobility management protocol, is dependent on a local mobility anchor (LMA) to process all the data and control traffics. Therefore, it has serious problems such as the tremendous traffic concentration into the core network and the triangle routing that causes inefficient and non-optimized routing path. In this paper, therefore, in order to alleviate these drawbacks, we propose a PMIPv6-based auxiliary distributed mobility management scheme considering each mobile node (MN)'s traffic locality. Performance evaluation results indicate that in most cases, except for when the MN's mobility rate is relatively very higher than the traffic rate, the proposed scheme shows better performance result than that of PMIPv6. Besides, it is demonstrated that the proposed scheme can be an effective alternative that can distribute significant loads on the LMA of the core networks to the MAGs of the edge networks.

Keywords: PMIPv6 · Traffic locality · Distributed mobility management

1 Introduction

PMIPv6 [1] relies on a central mobility entity (i.e., LMA) that manages both data plane and control plane for supporting mobility services. Therefore, it results in significant amount of data and control traffic being concentrated into the LMA of the core network, which may lead to serious bottleneck on the LMA and the triangle routing problem that causes inefficient and non-optimized routing path. Besides, the use of a central mobility entity such as LMA may be vulnerable to a single point of failure and degrade overall system performance [2–4]. In order to solve these drawbacks of centralized mobility management protocols such as MIPv6 and PMIPv6, the research issues on the distributed mobility management has been recently discussed in the IETF, which can be classified into the following two categories: the *partially distributed approach* where only data plane is distributed and the *fully distributed approach* where both data and control planes are distributed in the network.

Generally, each MN may have a variety of mobility/traffic characteristics. However, the current mobility management standards such as MIPv6 and PMIPv6 do not

Funding for this paper was provided by Namseoul University.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_163

consider each MN's mobility/traffic pattern. So, they may be inefficient for supporting explosively increasing MNs. Besides, the traffic load concentration into the central mobility entity and non-optimized routing path due to such a static and globally applicable approach of current mobility management standards may cause significant performance degradation throughout the networks. Therefore, it is preferable to design more efficient and more optimized per-user mobility management scheme considering each MN's mobility/traffic characteristics. In this paper, in order to alleviate these drawbacks of PMIPv6, we propose a per-user-based auxiliary distributed mobility management scheme for PMIPv6 networks (hereafter, we call it APMIPv6), which considers each MN's traffic locality.

2 The Proposed Scheme: Auxiliary PMIPv6 (APMIPv6)

The basic idea behind our proposed scheme is based on the following observations: most traffic destined to the MN tends to originate from a few of specific CNs which frequently communicate with the MN or relatively tremendous traffic volume originates from [5, 6]. Therefore, in this paper, for each MN, we define the CN-side mobile access gateway (MAG) (i.e., CN-MAG) from which the communications with the MN frequently originate or the heavy traffic originates as *frequently communicating MAG (FMAG)*. The proposed scheme considers each MN's traffic locality to efficiently distribute traffic load on the LMA to the MAGs located in the edge networks.

The brief description of the proposed APMIPv6 is as follows. For binding update, unlike PMIPv6, whenever the MN crosses the MAG coverage area, the new MN-side MAG (i.e., MN-MAG) sends the Proxy Binding Update (PBU) messages to the MN's FMAGs as well as its LMA. However, in this case, if the MN-MAG figures out that the MN does not have its FMAGs, it performs the binding update procedure just like in PMIPv6. Note that by checking the *FMAG address table* stored in the MN-MAG, the MN-MAG can determine whether the serving MN has its FMAGs or not and which MAGs are that MN's FMAGs. On the other hand, for packet delivery, if the CN-MAG receives data packets destined to the MN from the CN, it checks its *binding table* to see if it is the MN's FMAG or not. If the record is found at the binding table stored in the CN-MAG, it can obtain the MN-MAG's proxy CoA (care-of-address). Then, it directly sends the data packets to that address, not via the LMA (In this case, the packets can be transmitted through the optimized route path, and inefficient triangle routing can be avoided.). Otherwise, the CN-MAG performs the packet delivery procedure just like in PMIPv6 (i.e., it sends the data packets to the LMA).

In the following, more details of APMIPv6 are described based on the illustrative example shown in Figs. 1 and 2. Similar to PMIPv6, when the MN1 first enters the PMIPv6 domain and attaches to an access network connected to the MAG1 (**Step 1**), the MAG1 sends an AAA query message including the MN1's identifier and the MN1's current MAG's proxy-CoA (i.e., MAG1's proxy CoA) to the AAA server (**Step 2**). And then the MAG1 receives the AAA reply message from the AAA server (**Step 3**). Then, the MN-MAG (i.e., MAG1) checks its FMAG address table to see if the serving MN has its FMAGs or not. If the record for the serving MN is found at the FMAG address table, the MAG1 sends the PBU messages to the MAG3 (i.e., the

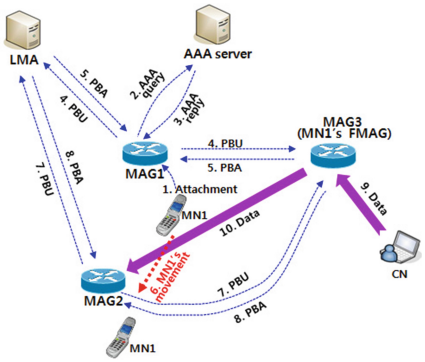


Fig. 1. The operation of APMIPv6

(a) Common FMAG address table stored in all the MAGs in the domain

ID	Each MN's FMAGs' Address List
MN1-ID	MAG3 proxy-CoA
MN2-ID	MAG4 proxy-CoA
MN5-ID	MAG3 proxy-CoA
	MAG6 proxy-CoA
.....	.....

(b) MAG3's binding table (right after the Step 8)

ID	Src. addr	Dest. addr
MN1-ID	MAG3 proxy-CoA	MAG2 proxy-CoA
MN5-ID	MAG3 proxy-CoA	MAG8 proxy-CoA
MN7-ID	MAG3 proxy-CoA	MAG5 proxy-CoA
.....	.....	.....

Fig. 2. The tables stored in the MAGs, which are defined in APMIPv6

MN1's FMAG) as well as its LMA (See the MN1's record at the FMAG address table shown in Fig. 2(a).). Note here that if the record for the serving MN is not found at the FMAG address table, the MAG1 sends the PBU message to the LMA only just like in PMIPv6 (**Step 4**). After that, if the MAG3 (i.e., MN1's FMAG) receives the PBU message from the MAG1, it updates the destination address field value of the MN1's record into MAG1's proxy-CoA at its binding table stored in MAG3, and sends the Proxy Binding Acknowledgement (PBA) message to the MAG1. On the other hand, the whole procedure of PBU/PBA message exchange with the LMA is the same as that of PMIPv6 (**Step 5**). Then, if the MN1 moves and attaches to MAG2 coverage area (**Step 6**), the AAA query/reply message exchanges between MAG2 and AAA server are performed just like in PMIPv6. Then, just like the procedures in Steps 4 and 5, the new MN-MAG (i.e., MAG2) sends the PBU messages to the MAG3 (i.e., the MN1's FMAG) as well as its LMA and receives the PBA messages from the MAG3 and the LMA. Note here that if the MAG3 receives the PBU message from MAG2, it updates the destination address field value of the MN1's record into MAG2's proxy-CoA at its binding table and sends the PBA message to the MAG2 (**Steps 7 and 8**).

Now, the packet delivery procedures are as follows. As shown in Fig. 1, the CN sends data packets towards the MN1 (**Step 9**). Then, the MAG3 (i.e., CN-MAG) checks its binding table to see if it is the MN's FMAG or not. If the record for the MN is found at its binding table, it directly sends the data packets to the destination address (i.e., the MAG2's proxy-CoA), not via the LMA (See the destination address field value of the MN1's record at the binding table shown in Fig. 2(b)) (**Step 10**). Note here

that if the record for the MN is not found at the binding table (i.e., if the MAG3 is not the MN1's FMAG), the MAG3 sends the packets to the LMA just like in PMIPv6.

To estimate the FMAGs for a particular MN, its traffic pattern throughout the days or weeks can be observed over a long period of time. For example, whenever an MN receives incoming packets, the MN-MAG may check the source addresses of the incoming packets (i.e., CN-MAG's proxy-CoA) and count the number of all the received packets or all the session arrivals. Considering that PMIPv6 is adopted in mobile communication systems, such information can be easily collected. The various ways to obtain a good estimate about the FMAG for each MN may be possible [5, 6]. In this paper, for simplicity, we assume that the FMAG information on each MN is commonly preconfigured in the FMAG address tables stored in all the MAGs.

3 Analytical Modeling

In this section, we analytically derive the total costs of PMIPv6 and APMIPv6, respectively. For the analysis, we consider the binding update (BU) cost and packet delivery (PD) cost during an *inter-session arrival time*. Table 1 shows the parameters used for analytical modeling. Generally, the total cost of a mobility management scheme can be composed of BU cost and PD cost. Thus, based on the operational procedures of BU and PD in PMIPv6, each cost of PMIPv6 can be expressed as

$$C_{BU}^{PMIP} = 2\,E(N_C)\,d_{lm} \tag{4.1}$$

$$C_{PD}^{PMIP} = 2\,E(L_S)\,(d_{lm} + \delta) \tag{4.2}$$

$$C_{Total}^{PMIP} = C_{BU}^{PMIP} + C_{PD}^{PMIP} \tag{4.3}$$

Table 1. Parameters used for analytical modeling

Parameter	Description
$E(L_S)$	The average session length (in number of packets)
$E(N_C)$	The average number of the MN's MAG coverage area crossings during an inter-session arrival time
d_{lm}	The transmission cost of a single packet between LMA and MAG
d_{cm}	The transmission cost of a single packet between CN-MAG and MN-MAG
d_{fm}	The transmission cost of a single packet between FMAG and MN-MAG
δ	The transmission cost of a single packet in the wireless link between MAG and MN
q	The ratio of the number of data packets destined to the MN originating from its FMAG to the total number of data packets destined to the MN
m	The average number of the MN's FMAGs in a domain
C_{BU}^{X}	Binding update cost of X scheme (X $\in$ {PMIP, APMIP})
C_{PD}^{X}	Packet delivery cost of X scheme (X $\in$ {PMIP, APMIP})
C_{Total}^{X}	Total mobility management cost of X scheme (X $\in$ {PMIP, APMIP})

Similarly, the BU cost, the PD cost, and the total cost of APMIPv6 can be calculated as

$$C_{BU}^{APMIP} = 2E(N_C)\,d_{lm} + 2mE(N_C)\,d_{fm} \tag{4.4}$$

$$C_{PD}^{APMIP} = qE(L_S)(2\delta + d_{cm}) + (1-q)(2E(L_S)(d_{lm} + \delta)) \tag{4.5}$$

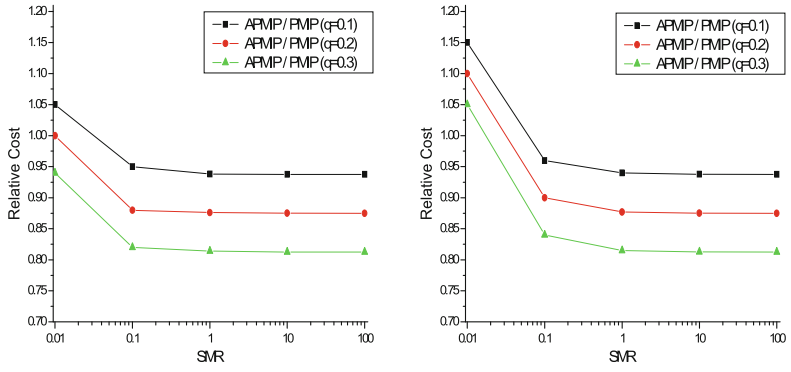
$$C_{Total}^{APMIP} = C_{BU}^{APMIP} + C_{PD}^{APMIP} \tag{4.6}$$

where in Eq. (4.4), $2mE(N_C)\,d_{fm}$ means the binding update cost that the MN-MAG registers with m FMAGs. On the other hand, in Eq. (4.5), $E(L_S)(2\delta + d_{cm})$ means the packet delivery cost in case the CN-MAG is the MN's FMAG, and $2E(L_S)(d_{lm} + \delta)$ means the packet delivery cost in case the CN-MAG is not the MN's FMAG (i.e., in this case, the packets are routed indirectly through the LMA), respectively.

4 Numerical Results

In this section, we conduct the performance analysis based on each cost of PMIPv6 and APMIPv6 derived from Sect. 3. For numerical analysis, we set parameter values as $d_{lm} = 30$, $d_{cm} = d_{fm} = 20$, and $\delta = 3$, respectively. These values were assigned based on the work in [7] by considering the hop distance between mobility agents (i.e., LMA and MAG).

Figure 3(a) and (b) show the effects of SMR (Session-to-Mobility Ratio) on the total costs of PMIPv6 and APMIPv6 when *E(Ls)* is set to 500 (in number of packets) and m is set to 1 and 2, respectively. For the analysis, we consider *the relative total cost of APMIPv6* when the total cost of PMIPv6 is normalized to 1. When SMR is very small (e.g., in case of SMR = 0.01, in other words, MN's session arrival rate: MN's MAG coverage area crossing rate = 1 : 100), the cost of PMIPv6 shows slightly better



(a) *E(Ls)* = 500 and *m*=1 (b) *E(Ls)* = 500 and *m*=2

Fig. 3. Effects of SMR, q and m on the total costs of PMIPv6 and APMIPv6

performance than that of APMIPv6. This case corresponds to the exceptional scenario where the MN is highly mobile almost without any communication with the CN. So, the benefit of APMIPv6 cannot be obtained in this case. However, in all the cases except for this one (e.g., in cases of SMR = 0.1, 1, 10 and 100, respectively), as SMR gets larger, the cost of APMIPv6 shows better performance than that of PMIPv6. Besides, such phenomenon gets prominent as q gets larger. This is due to the following facts: as SMR gets larger, the packet delivery cost gets dominated. So, the benefit of APMIPv6 gets prominent because of sending data packets through the optimized route path, not via the LMA in case that the CN-MAG is the MN's FMAG. Also, the larger value of q means that the more data packets are sent from the FMAG through the optimized route path, not via the LMA. Thus, as shown in Fig. 3(a) and (b), the performance of APMIPv6 gets more superior as q gets larger.

5 Conclusions and Future Works

In this paper, inspired by the observations and facts that most MNs, to some extent, has traffic locality, we proposed an efficient PMIPv6-based auxiliary mobility management scheme for distributing the loads on the central mobility anchor point, that is LMA, to the MAGs located in the edge networks. Numerical results demonstrated that the proposed APMIPv6 shows the apparent potentials to mitigate the serious drawbacks of PMIPv6 (e.g., LMA bottleneck problem and triangle routing, etc.) while reducing the total mobility management cost. Future research directions will be focused on the in-depth work on the MN's traffic characteristics and the determination of the FMAGs on each MN because the appropriate configuration of the FMAGs on each MN is crucial factor for the performance of APMIPv6. The research on the adaptive configuration of the FMAGs on each MN will be also conducted.

References

1. Gundavelli, S., et al.: Proxy Mobile IPv6, IETF RFC 5213 (2008)
2. Yokota, H., Seite, P.: Use Case Scenarios for Distributed Mobility Management. IETF Internet-Draft, draft-yokota-dmm-scenario-00.txt (2010). KDDI Lab
3. Koh, S., et al.: Use of Proxy Mobile IPv6 for Distributed Mobility Control. IETF Internet-Draft, draft-sjkoh-mext-pmip-dmc-03.txt (2011)
4. Jung, H., et al.: Distributed mobility control in proxy mobile IPv6 networks. IEICE Trans. Commun. **F94-B**, 2216–2224 (2011)
5. Kong, K.-S.: Performance analysis of profile-based location caching with fixed local anchor for next-generation wireless networks. IEICE Trans. Commun. **E91-B**(11), 3595–3607 (2008)
6. Shivakumar, N., et al.: Per-user profile replication in mobile environments: algorithms, analysis, and simulation results. ACM Monet **2**(2), 129–140 (1997)
7. Kong, K.–S., et al.: Mobility management for all-ip mobile networks: mobile IPv6 vs. proxy mobile IPv6. IEEE Wireless Commun. **15**, 36–45 (2008)

A Study on Worker's Positional Management and Security Reinforcement Scheme in Smart Factory Using Industry 4.0-Based Bluetooth Beacons

SangIl Park and SeoukJoo Lee(✉)

Graduate School of Computer and Information Technology, Korea University, Anam Campus, Anam-dong 5-ga, Seongbuk-gu, Seoul, Korea
{catzero, seouklee}@korea.ac.kr

Abstract. The industrial accidents in the factories are continuously occurring globally and accordingly, the property damages and the number of human casualties are also increasing. The potential for the accidents at chemical plants are especially high as these factories heavily use or store many kinds of hazardous substances. Thus, separate preventive measures or tools are required. This study focused on a system that can prevent accidents and reinforce the security in a Smart Factory. The system uses Industry 4.0-based Bluetooth beacons to determine a worker's position and enhance the security level. Converting to a Smart Factory, globally located factories are adding variety to all the processes in manufacturing and logistics. At the same time, some new and innovative methods are required to adapt to such changes, especially in the fields of safety and security. The existing communication method between beacons is that relevant application will be installed in the Bluetooth-activated devices such as Smart Pad, smart phone or other similar device to calculate the distance between the device and the beacon. The distance information will be then delivered to the server to control other devices/equipments in the factory. However, this method has a problem of using the Smart Pads in the factory due to the security or spatial characteristics of the factory and the system will cease to operate when battery life expires. To deal with such problems, a beacon has been attached on the worker's safety helmet and synchronize beacon's ID with worker's ID so that the system can check the both signals to control worker's access to the factory or let the security to take measures. Thus, this system allows an effective worker access control, notifications of entering danger zones, establishment of an efficient working condition through evaluating worker's movements, estimating the positions of workers in the event of accident, and safe log-ins for factory's security. Just by attaching the beacon on the worker's helmet, no additional application installations, devices, and external network or GPS system will be needed so that an independent network can be established at the factories distant from cities.

Keywords: IoT · Smart Factory · Bluetooth · BLE · Safe environment

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_164

1 Introduction

Recently, seven workers got hurt at the company located in the S city in the Republic of Korea (ROK). Their insensitivity towards safety has been blamed before when a worker was killed in a fall 15 days ago. In India, thousands of people died from the world's worst chemical factory accident in Dec. 1984. Also, in China last year, toxic substances leaked from the chemical factory in the Shandong province, bringing much damages to the factory workers and nearby residents. These incidents are just a few examples around the world and they never seem to be reducing. Especially, being a highly technology-intensive process industry, chemical factories handle massive amount of hazardous chemicals so that there always is a possibility of an accident. Therefore, a reliable preventive system is essential. This system uses Industry 4.0-based Bluetooth beacons to determine a worker's position and enhance the security level [1–6].

2 Related Study

The Smart Factory is a part of the Industry 4.0 strategy where the ITC technology has been combined with the manufacturing sector to enable an automated production system through simulations. The production and service elements involved in this future production system are connected by the network and exchange information to achieve optimized production and self-controlled process. Accordingly a safety management system that meets such a Smart Factory must be studied first [7–11].

The ultimate goal of the Smart Factory is to combine 4M1E elements (i.e., Man, Machine, Material, Method and Energy) to deduce an integrated synergy effect. That is, automatically producing the products with minimum raw materials, lowest amount of energy, and best possible method available so that the rate of defectives should be low but the production yield will be greater. Also, the system requires small manpower to effectively produce the products. The Smart Factory collects the on-site 4M1E information in real time and provide them to the management to let them make the best decisions. At the same time, the system informs customers about the expected delivery dates while notifying the present factory conditions to the factory manager in real time. The factories around the world are bringing the changes to every step of the production and distribution processes while they are attempting to convert their factories into the Smart Factories. The workers' safety and efficiency must be considered during the conversion process since a large amount of products will be produced with a minimum number of workers. Also, for the technology-intensive manufacturing industry, production technology, safety and system security must be considered first [11–15].

3 A Study on Worker's Positional Management and Security Reinforcement Scheme in Smart Factory

Figure 1 shows a Bluetooth Beacon System configuration and the location of Wi-Fi device following the structure of the factory. There are a number of smart devices/equipments inside the factory to control facilities, manage raw materials, stocks,

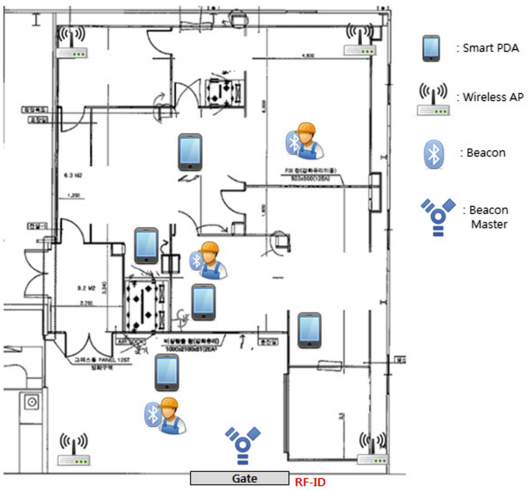


Fig. 1. A bluetooth beacon system configuration and the location of Wi-Fi device following the structure of the factory

products, and understand the current situation of materials supply and demand, as well as factory production status. The smart equipments compute the positions of beacons by using the Bluetooth signals. In order to calculate the exact position, the distances between the smart equipments must be estimated. The error range in calculation of distance is about 3 m and if there are no obstacles in the space, the measurement can be achieved as far as 50 m to 70, or 30 m to 30 m when there are any obstacles. Estimate the worker's exact distance by calculating the distance between each smart equipment.

The signals from the beacon and the smart equipment will be stored at the central sever every three seconds so that if a worker is positioned in the restricted zone, a warning SMS/E-mail will be sent to the worker and the safety manager. When a factory worker attempts to log on to the smart equipments which control and manage factory facilities, signal IDs will be compared through the server and the only equipments having the beacon distance of less than 3 m will be accepted. If the worker deviates the distance over 3 m, he will be logged out, preventing others to control the equipment(s). By evaluating the movements in real time, the most efficient traffic route suitable to the factory environment can be deduced.

Also, in the event of accident, the factory manager can concentrate on prompt life saving measures and accident controls by determining exact positions of workers and their number within the factory. Figure 2 is an algorithm used for the security validation and warning system. When a worker accesses the factory, the condition of his/her beacon will be checked (e.g., beacon's voltage level, beacon ID and RF-ID for any mismatches, beacon signal strength). If the beacon is normal access time will be sent to the server and if there are any problems in the beacon, the server will make decisions depending on the type problem occurred. The current position of the worker is continually sent to the server through smart device (PAD).

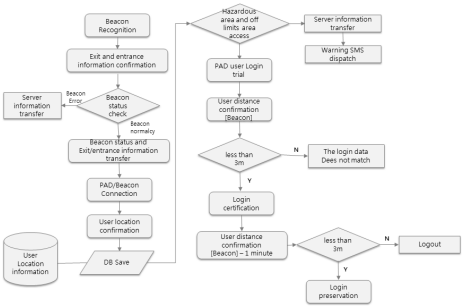


Fig. 2. Algorithm for worker's security validation and warning mechanism

The beacons are to be connected to a multiple number of smart devices and a beacon manager to provide the distance information and access data to the server. The position data will be accumulated at the server to create a statistics to determine the most efficient traffic route and environment in the factory.

If the workers approach or enter the restricted or unauthorized areas, a warning SMS will be sent to the worker and safety manger simultaneously with alarms.

When a worker attempts to log on to smart equipment used for the production, the distance between the worker and the beacon will be measured. If the distance remains within 3 m, he will be accepted and if any communications are not established between them, the smart equipment will continuously measure the distance every one minute. In this case, if the distance exceeds 5 m, the smart equipment will be shut off automatically or stay logged in if the distance is within the range of 5 m.

Figure 3 shows entire system configuration of a factory where the chemicals for the semiconductors production are being produced. The manufactured chemicals will be delivered to the customers with double-packed containers. Its primary process includes solution-mixing and concentration control processes. The final product is produced by filtering the premixed and controlled solution through 3 to 5 filters. Also, the final products will be primarily packed in the customized container with labels and then wrapped with the sun-blocking film before putting it in the box. The boxed products are then tied together on a pallet and stored in the warehouse. Sometimes the products are released on pallets but otherwise, they will be re-packed following the orders by the customers. The production line will be cleaned after production process is complete using cleaning solutions to get ready for the later production.

The number of raw materials is often small but in order to maintain the quality, exact proportion is required along with good inventory control. The raw materials should be controlled by distinguishing them by their minimum packing units. The quality inspection process is divided into three steps: raw material inspections, intermediate product inspection and final product inspection. The raw material inspection literally inspects the materials to be used and in the intermediate inspection step, the accuracy of mixing and concentration will be checked. In the final step, the product will be checked finally after the filtering process. All the inspection data will be stored with the product IDs.

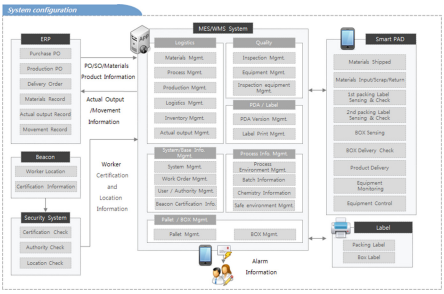


Fig. 3. A whole system configuration

In general, the production facilities management system is operated through the touch-operated process management display or PC connected to the PLC network. The display only shows the information about a unit facility so that development, operation, and maintenance are not easy to perform. When showing the production management system on a process PC, there are several problems in the factories where production site is too small or sensitive to the pollutants. Therefore, the factories introduced a Smart PAD which can deal with such problems and easy to move wearing a dust-prove garment (Figs. 4 and 5).

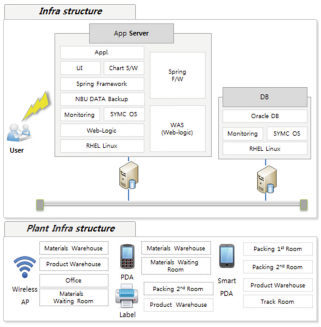


Fig. 4. Infra structure of factory

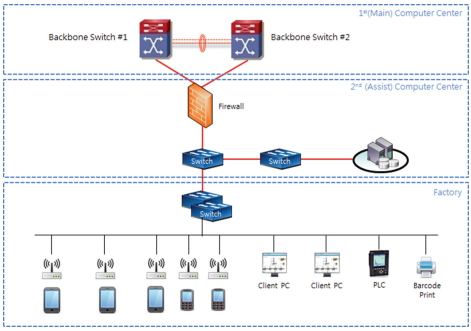


Fig. 5. Process network configuration

The Bluetooth beacons-based Smart Factory security management system proposed in this study has been constructed with the Java spring and is divided into a Web UI for MES/WMS and an Android UI for Smart PAD. The production information are stored in the Oracle RDBMS together with worker's positional data.

For the security of the factory, interior of the factory will be covered by a separate internal network (Factory Area/FA network) where the FA network would not be accessed externally. All the wireless equipments are to be subjected to the security settings (i.e., registering their MAC addresses to get an access to the FA network). Also, all the equipments including Smart PAD, Client PC, PLC devices, PDA, and barcode printer should be connected through the FA network. The computer room is first connected to the factory's network through Switch and the firewall, and then to the

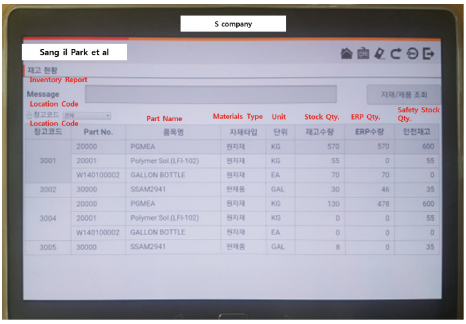


Fig. 6. Taking an inventory of the factory using smart PAD

office network (Office Area/OA network). The reason for this network separations is to physically prevent external influences which may be harmful to the factory facilities and information networks, as well as to increase the network speed. An example is presented in Fig. 6 below.

4 Conclusion and Future Work

The factory automation through Smart Factory is being introduced in several countries as a preferable means of increasing the efficiencies in the manufacturing sector and generating the higher value-added businesses. In this Smart Factory, the human resources management is one of the most important elements, along with machines, materials and methods used so that preparing some special safety measures are essential in promoting an efficiency among workers.

This study focused on the worker's positional information management and systemic validation method which utilizes the Bluetooth-based beacons to secure safe factory environment and security. With this system, establishment of a safety environment and security measures within the Smart Factory including activities such as controlling the workers' factory accesses, issuing a warning when they approach restricted or hazardous areas, determining the positions of workers in the factory in the event of accidents can be achieved. Just by attaching a beacon to a worker and without any separate applications, this system can be applied to the factory. Moreover, no external networks or GPS systems are needed for this system so that it is possible to construct an independent network in the factories distant from urban areas.

The ultimate goal of the Smart Factory is creating an integrated synergy effect through establishing connections between elements of 4M1E and promoting automation process. That is, the system should be able to produce good products automatically, effectively and steadily with lesser people, minimum amount of raw materials/energies, and the best possible means and yields. By applying the beacon-based worker positioning system proposed in this study onto the facilities in the Smart Factory, the factory manager will be able to assign a worker for an inspection duty on a specific time and inform him/her of the best position where he/she can move

most efficiently to fulfill his/her duty. We expect that the system will be also useful in determining an adequate number of workers for a specific production process. Some comparative analyses will be conducted for the UML and the entire framework used in this study against other studies in our future journal paper, along with the studies related to the mode changes that will surmount the drawbacks in the Bluetooth technology, such as the shadow areas and signal attenuations, for example.

References

1. Eagle, N., Pentland, A.: Social serendipity: mobilizing social software. In: IEEE Pervasive Computing 4.2, pp. 28–34 (2005)
2. Lu, M., Chen, W., Shen, X., Lam, H.C., Liu, J.: Positioning and tracking construction vehicles in highly dense urban areas and building construction sites. In: Automation in Construction 16.5, pp. 647–656 (2007)
3. Zhu, J., Zeng, K., Kim, K.H., Mohapatra, P.: Improving crowd-sourced Wi-Fi localization systems using bluetooth beacons. In: 2012 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON), pp. 290–298. IEEE (2012)
4. Chang, Y.J., Chu, Y.Y., Chen, C.N., Wang, T.Y.: Mobile computing for indoor wayfinding based on bluetooth sensors for individuals with cognitive impairments. In: 2008 3rd International Symposium on Wireless Pervasive Computing, ISWPC 2008, pp. 623–627. IEEE (2008)
5. Kajioka, S., Mori, T., Uchiya, T., Takumi, I., Matsuo, H.: Experiment of indoor position presumption based on RSSI of bluetooth LE beacon. In: 2014 IEEE 3rd Global Conference on Consumer Electronics (GCCE), pp. 337–339. IEEE (2014)
6. Huh, J.H., Yohan, B., Seo, K.: Bluetooth-tracing RSSI sampling method as basic technology of indoor localization for smart homes. Int. J. Smart Home **10**(10), 1–14 (2016)
7. Huang, H., Gartner, G., Schmidt, M., Li, Y.: Smart environment for ubiquitous indoor navigation. In: 2009 International Conference on New Trends in Information and Service Science, NISS 2009, pp. 176–180. IEEE (2009)
8. Kalia, M., Garg, S., Shorey, R.: Efficient policies for increasing capacity in bluetooth: an indoor pico-cellular wireless system. In: 2000 IEEE 51st Vehicular Technology Conference Proceedings, VTC 2000-Spring Tokyo, vol. 2, pp. 907–911. IEEE (2000)
9. Chawathe, S.S.: Low-latency indoor localization using bluetooth beacons. In: 2009 12th International IEEE Conference on Intelligent Transportation Systems, pp. 1–7. IEEE (2009)
10. Herrera, M.M., Bonastre, A., Capella, J.V.: Performance study of non-beaconed and beacon-enabled modes in IEEE 802.15. 4 under bluetooth interference. In: 2008 The Second International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies, UBICOMM 2008, pp. 144–149. IEEE (2008)
11. Huh, J.H., Seo, K.: RUDP design and implementation using OPNET simulation. In: Park, J. J.(J.-H.)., Stojmenovic, I., Jeong, H.Y., Yi, G. (eds.) Computer Science and Its Applications. LNEE, CUTE 2014, vol. 330, pp. 913–919. Springer, Heidelberg (2015)
12. Huh, J.H., Lee, D.G., Seo, K.: Design and implementation of the basic technology for realtime smart metering system using power line communication for smart grid. In: Park, D.-S., Chao, H.-C., Jeong, Y.-S., Park, J.J.(J.-H.) (eds.) Advances in Computer Science and Ubiquitous Computing. LNEE, CUTE 2015, vol. 373, pp. 663–669. Springer, Singapore (2015)

13. Huh, J.H., Otgonchimeg, S., Seo, K.: Advanced metering infrastructure design and test bed experiment using intelligent agents: focusing on the PLC network base technology for Smart Grid system. J. Supercomput. **72**(5), 1862–1877 (2016). Springer, USA
14. Park, J.H.: Design of the real-time mobile push system for implementation of the shipboard smart working. In: Park, D.-S., Chao, H.-C., Jeong, Y.-S., Park, J.J.(J.-H.). (eds.) Advances in Computer Science and Ubiquitous Computing. LNEE, CUTE 2015, vol. 373, pp. 541–548. Springer, Singapore (2015)
15. Inoue, Y., Sashima, A., Ikeda, T., Kurumatani, K.: Indoor emergency evacuation service on autonomous navigation system using mobile phone. In: 2008 Second International Symposium on Universal Communication, ISUC 2008, pp. 79–85. IEEE (2008)

A Study of the Extended Definition of Relation for Research Content Based Traceability

Jong-Won Ko, Jae-Young Choi(✉), and Young-Hwa Cho

College of Information and Communication Engineering,
SungKyunKwan University, Suwon-si, South Korea
{jwko0820, jaeychoi, choyh2285}@skku.edu

Abstract. Trace artifacts refer to any residual data or marks of the R&D process that are made amenable to being trace. we applied this term to Research Contents for R&D project in our previous research. It is called Research Descriptor also known as RD. Furthermore, the existing researches failed to fully consider the research items existing in the form of various document files and SW configuration items, for example, the support for traceability among SW UML models and source codes, UI Form and test cases and issues on traceability with the conducted R&D process to make such outcomes. In this paper, we proposed extended definitions of various perspective relationship which to support traceability for research documents, processes and source codes during the whole period of the R&D project. Also, we considered definition of logical relationship between research contents and discussed how to manage the R&D project by tracking the RD relation type.

Keywords: Research content relation · Research Descriptor · Research Descriptor Based Traceability

1 Introduction

Trace artifact is one of the trace elements and is qualified as either a source artifact or as a target artifact when it participates in a trace. The size of the traceable unit of data defines the granularity of the related trace. Trace artifacts refer to any residual data or marks of the R&D process that are made amenable to being trace [1]. we applied this term to Research Contents for R&D project in our previous research. It is called Research Descriptor also known as RD. Furthermore, the existing researches failed to fully consider the research items existing in the form of various document files and SW configuration items, for example, the support for traceability among SW UML models and source codes, UI Form and test cases and issues on traceability with the conducted R&D process to make such outcomes [2, 3].

Therefore, in our previous work, we defined the scope and target of traceability regarding the Research Descriptor (RD) traceability and we also defined RD based Traceability Information Model. The Research Descriptor is the research contents in various forms produced over the whole period of the R&D project including the execution process of the R&D project and keywords and key sentences of the outputs.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_165

In this paper, we proposed extended definitions of various perspective relationship which to support traceability for research documents, processes and source codes during the whole period of the R&D project.

Also, we considered definition of logical relationship between research contents and discussed how to manage the R&D project by tracking the RD relation type.

Section 2 describes the existing researches on RD based traceability and RD based Traceability Information Model and Sect. 3 covers the information on the extended definition of relation between research contents for Traceability and describes an example of extended RD relation for Traceability. Section 4 gives the conclusion of the proposal of this paper and future research plans.

2 Related Works

This chapter describes the key concept in the Research Descriptor Based Traceability from four perspectives of traceability and also describes how the information structure is composed and what kind of relation type the information has, and the Research Descriptor based Traceability Information Model.

2.1 Research Descriptor Based Traceability

The Research Descriptor generated in the R&D project has traceability to support the semantic analysis technology, similarity analysis, or semantic-based tests. The traceability for Research Descriptor can be defined as follows from four perspectives [4].

- Traceability from the perspective of work products: Traceability between Document RD which is the result in the form of document files Tracing of various versions for one document file and traceability for other related document files.
- Traceability from the perspective of process: Tracing of relation between the generated Document RD and any tasks or activities on the project WBS.
- Traceability from the perspective of the monitoring quality metrics: Tracing of Document RD that affects monitoring quality metrics.
- Traceability from the perspective of software development: Tracing of components or class related to software development associated with generated Research Descriptor or source files.

2.2 Research Descriptor Based Traceability Information Model

As reviewed in the relevant researches in Sect. 2.1, what is the information to be traced, how to define the information structures and how to set the relation between the information to be traced will be explained using the TIM model when defining the traceability through the TIM model. Therefore, in our previous work, we also defined the TIM model as shown in Fig. 1 to define RD based traceability in executing the R&D project [4].

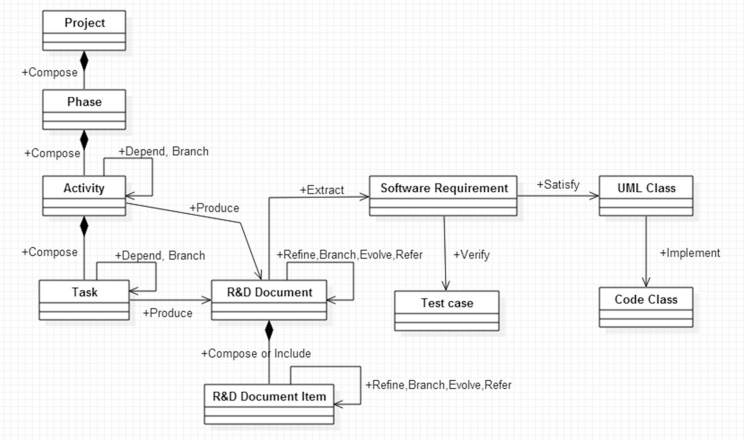


Fig. 1. Research descriptor based traceability information model

Figure 1 demonstrates the structure and the Compose relation between Project, Phase, Activity and Task from the process aspect among the targets to be traced first, and the Produce relation with R&D Documents produced through certain Activity and Task. Also, the relation between Activity and Activity, Task and Task can be expressed as the Depend (FS, SS, FF, SF) and Branch relation.

Also, it defined various forms of outputs produced over the entire period of the R&D Project as R&D Document; described the Section or Subsection within the Document as R&D Document Item; and defined the relation between Section and Subsection as Compose and the relation between Document and Document and Document Item and Document Item as Refine, Depend, Branch, Evolve and Refer. Each relation type will be explained in detail in the next Section.

In addition, the relations between Software Requirement or Test case and UML Class or Code Class will be expanded into Extract or Satisfy, Verify and Implement as a target to support traceability from the SW perspective.

3 The Extended Definition of Relation for Research Content Based Traceability

3.1 The Extended Definition of Relation Type for Research Descriptor

As reviewed Sect. 2, in order to support definition of various RD relation between research contents we focused on extended definition of RD relation types from our previous works. The relation type was redefined as follows depending on each identified relation type. To define extended RD relation we mainly focus on two relation types below.

– Artifact RD Relation
  : Relation between document and document, document and code, code and code.

– Process RD Relation
: Relation between process and process, process and document, code and code.

Figure 2 shows Definition of the relations for Artifact RD (Document and Code) and Process RD. and we also defined the extended RD relation by Tables 1, 2 and 3.

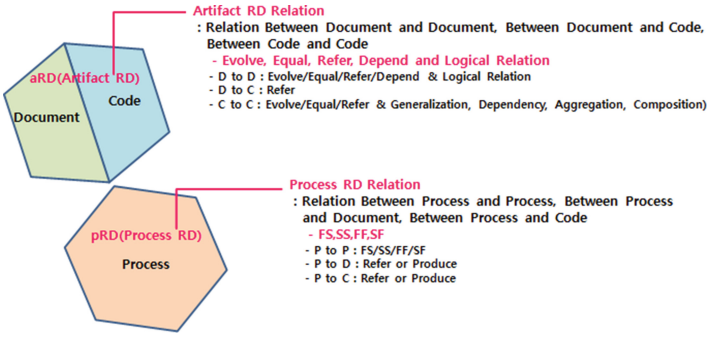


Fig. 2. Definition of relation for artifact RD and process RD

Table 1. Definition of RD relation for document and code

Relation	Description	Implementation for relation verification
Evolve	When the target RD add new or detailed contents from source RD and these RDs have a same topic, there should be high similarities between RDs	1. bool is_diff(string str1, string str2) 2. bool is_same_toic (string str1, string str2) 2.1 string parser_root (string str) 2.2 int[] topic_vec(string str) 2.3 double sim_cosine (int[] var1, int[] var2)
Equal	When target and source RD contents are exactly same or semantically same	1. bool is_diff(string str1, string str2) 2. string parser_root (string str) 3. int[] text2vector(string str) 4. double sim_cosine(int [] var1, int[] var2)
Refer	When the target RD refers to the internal and external research contents of the R&D project, researcher can set the Refer relation	1. string parser_NN (string str) 2. string keyword(string str, int count) 3. double sim_jaccard (int[] var1, int[] var2)

Table 2. Definition of logical relation between research contents

RD relation	Description
Top and bottom	The relationship between subordinates and superiors
Amplification	Expatiate on relationship
illustration	An illustration is an example or a story which is used to make a point clear
Summary	A Summary of something is a short account of it, which gives the main points but not the details
Cause and effect	If there is a causal relationship between two RDs, one RD is responsible for causing the other RD
Addition	Additional RDs are extra RDs apart from the ones already present
Sequence	A sequence is a number of RDs that come one after another in a particular order

Table 3. Classification of document-code-process RD relation definition

D-C-P type	RD relation	Remarks	RD type
D to D	Evolve, Equal, Refer Logical relation	Need to consider section and subsection in document	Artifact RD (aRD)
D to C	Refer		
C to C	Evolve, Equal, Refer Generalization, Dependency, Aggregation, Composition: UML class relation	Need to consider code granularity (Class or Method)	
P to P	FS, SS, FF, SF	Process dependency	Process RD (pRD)
P to D	Refer, Produce	Process outputs as documents and codes	
P to C	Refer, Produce		

3.2 Example of Extended RD Relation for Traceability

Figure 3 demonstrates an example of extended RD relations between research contents as sections in research documents. Shown as Fig. 3, It has Evolve relation between research contents and logical relation as Sequence and Summary relation.

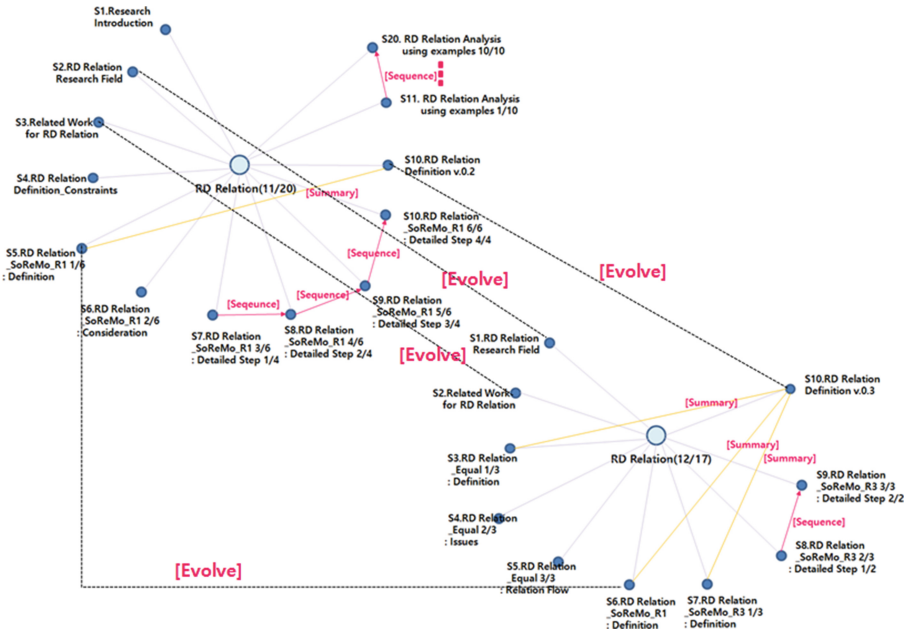


Fig. 3. An example of extended RD relation between research contents for traceability

4 Conclusion and Further Works

As reviewed in Sect. 3, this paper defined the extended Research Descriptor (RD) relation for research contents based traceability and RD based Traceability Information Model. And we also applied the extended RD relation types to support RD based Traceability that was defined in our previous works. It considered the targets and perspectives that require traceability and the tracking relation types to support traceability in executing the R&D project from various aspects, and suggested a research on traceability linked to the quality of R&D project through the correlation between traceability based on various RD relation types.

Going forward, further research is scheduled on the existing researches on similarity to verify the extended RD relation types to support traceability that was defined in this paper and implementation of RD based Traceability Framework that supports the R&D project.

Acknowledgements. This research was supported by the Next-Generation Information Computing Development Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2014M3C4A7030503).

References

1. Gotel, O., Cleland-Huang, J., Hayes, J., Zisman, A., et al.: Traceability fundamentals. In: Software and Systems Traceability, pp. 3–22 (2012)
2. Alobaidi, M., Mahmood, K.: Semantic approach for traceability link recovery using uniform resource identifier (STURI). In: International Conference of Software Engineering, Research and Practice (SERP) (2015)
3. Lucia, A., Fasano, F., Oliveto, R., Tortora, G.: Recovering traceability links in software artifact management systems using information retrieval methods. ACM Trans. Softw. Eng. Methodol. **16**(4), 13:1–13:50 (2007). Article 13
4. Ko, J.-W., Kim, S.-T., Choi, J.-Y., Cho, Y.-H.: Contents based traceability between research artifact and process for R&D projects. In: Park, D.-S., Chao, H.-C., Jeong, Y.-S., Park, J.J.(J. H.) (eds.) Advances in Computer Science and Ubiquitous Computing. LNEE, vol. 373, pp. 891–898. Springer, Singapore (2015)
5. Vingate, L.M.: Project Management for Research and Development: Guiding Innovation for Positive R&D Outcome. CRC Press, Boca Raton (2015)

Transforming Algorithm of 3D Model Data into G-code for 3D Printers in Distributed Systems

Sungsuk Kim[1] and Sun Ok Yang[2(✉)]

[1] Department of Computer Science, Seokyeong University, Seoul, South Korea
sskim03@skuniv.ac.kr
[2] Department of Computer Science, Gachon University, Seongnam, South Korea
soyang9149@gmail.com

Abstract. 3D printing is a process of making 3D solid object from a digital STL file. 3D printer operates according to G-code. A STL file is composed of lots of facets, and the number increases depending on the size or precision of 3D model. In this paper, we have interests on devising an algorithm to transform STL file into G-code in distributed systems. The algorithm is divided into two steps: first, grouping facets according to Z-axis value, and second, generating G-code from facets. Through the simulation works, we come to know that the transformation can be done well in distributed manner.

Keywords: 3D printer · 3D model · STL file · G-code Distributed computing

1 Introduction

3D printing is a process of making three dimensional solid objects from a digital file. The creation of a 3D printed object is achieved using additive processes, where an object is created by laying down successive layers of material until the object is created. 3D printing starts with making a virtual design of the object you want to create. The design is for instance, a CAD file. Recently, 3D scanner is another different example to generate a 3D model. When a 3D model is prepared, next process called *slicing*, is needed. Slicing is dividing a 3D model into hundreds or thousands of horizontal layers and needs to be done with software. G-code is used to control 3D printer; that is, it instructs where to move, how fast to move, and what path to move. Finally, the 3D model is ready to be printed in a 3D printer.

Generally, 3D model can be represented in STL format and transformed into G-code for 3D printer. A STL file is composed of lots of facets, which is a flat faces on geometric shapes. The facet number in an object highly depends on the precision or size of 3D model. For example, an object of 10 cm height has 7.5×10^2 facets, but if the height is 100 cm, the object is composed of 5×10^5 facets. In this case,

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF- 2015R1D1A1A01061173).

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_166

transformation time is needed about 10^3 times more. Of course, more time to translate is needed.

In this paper, we devise a new distributed transformation algorithm from STL file to G-code. The process is divided into two steps: first, grouping facets according to Z-axis value, and second, generating G-code from facets. In our devised algorithm, the steps will be done in distributed manner. To do so, there are n processing nodes and node N_0 plays a role of coordinator. Each step will be explained in next subsections in detail.

1.1 Step 1: Grouping Facets

Each facet can be represented as 3 coordinates (x, y, z) covering one small part on object surface. Slicing is generally processed according to Z-axis but there is no rule among facets stored in a STL file. Thus, as first step, z_{max} and z_{min} is extracted and stored separately for each facet and then all facets are split into several groups depending on z_{min}. For example, f_1 is a facet on a cylinder and the coordinates are like Fig. 1. From the facet, the greatest and the smallest Z-value is 9 and 4.

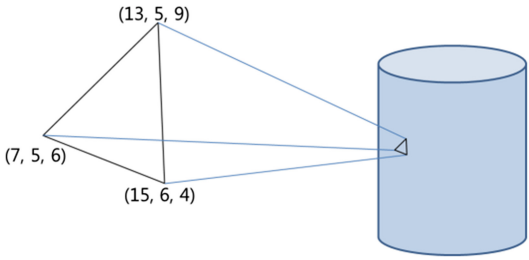


Fig. 1. An example of a facet f_1

Before grouping, preprocessing has to done at node N_0; for each facet, a unique ID is given and z_{max} and z_{min} is extracted and stored as a form of (ID, z_{max}, z_{min}) into F_{set}. (ID, z_{max}, z_{min}) will be hereafter called $FInfo$. All data in F_{set} is split into n parts equally where n is the number of processing nodes except node N_0. When N_0 hands over a part of F_{set} data to node N_i ($1 \leq i \leq n$), N_i sorts them by z_{min} in ascending order and finally returned the result to N_0.

1.2 Step 2: Generating G-code from Facets

When node N_0 gets the result from node N_i, it again divides them into n groups according to the Z-coordinator position. In this way, from all results, n group data are generated; it means that z_{min} of all facets in group 1 is smaller than z_{min} of any facet in group 2. For example in Fig. 2, node N_0 gets 4 results from the other nodes where all facets are sorted. It makes 4 groups by dividing facets according to the height of the 3D model (in Fig. 2, Group_1 to Group_4). Each Group_i data will be delivered to node i ($1 \leq i \leq 4$).

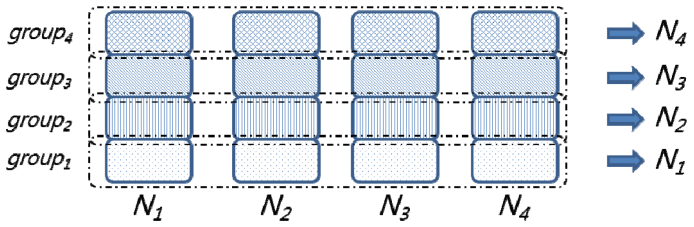


Fig. 2. After process N_0 gets back the result from the other nodes, it redistributes to them

When node i gets group data which contains facets sorted by z_{min}, it will transform facets to G-code. That is, for example, node 1 gets facets data of which Z value is from height 0 to height/4. Thus, it first sets the position of plate of 3D printer to 0. It (1) finds facets which intersect the plate, (2) calculates intersection points, and (3) converts G-code. Next, the plate position moves up by constant value (δ) and do the process (1) to (3) until plate position is height/4.

The following figure shows how to calculate the intersected position between a facet and the plate (Fig. 3).

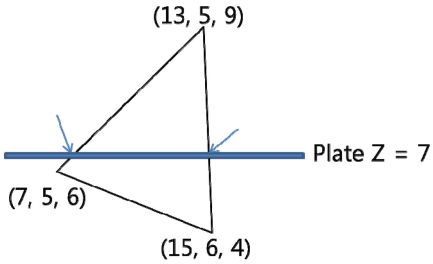


Fig. 3. Example: plate position is 7

1.3 The Algorithm

Let us assume that there are 5 nodes in a system. N_0 is a coordinator node and the other 4 nodes are slave. The algorithm operates as follows:

(1) **Node N_0:** it first opens a STL file, do preprocessing on it (that is, extracting *FInfo* data), divide it into 4 parts and delivers each part to a node N_i $(1 \leq i \leq 4)$.

(2) **Slave node N_i:** When a slave node gets part data, it just splits all *FInfos* in the data into 4 groups depending on Z_{min} value and returns the group data to N_0.

(3) **Node N_0:** When N_0 gets results from all slave nodes, it merges all groups i into $group_i$. And it redistributes $group_i$ to node i.

(4) **Slave node N_i:**

① Sets the plate position of 3D printer to the bottom;

② Finds faets which intersect the plate and calculates the intersection points;

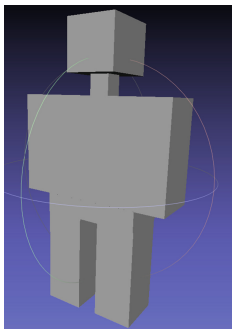
③ Converts them into G-code

④ Moves the plate up by δ; if the plate position is lower the top, go to ②.

2 Experiments and Results

In distributed systems, the number of processing nodes is a critical factor. And the transformation job highly depends on the size or precision of 3D digital object. In the experiments, we make a 3D model data.

We first check the number of facets. Since the used 3D model is very simple, it is composed of only 96 facets when the surrounding length is 5 cm. But the number grows 1749 when the length is 20 cm and transformation time changes 7 min to 107 min.



The next table is relates with transformation time depending on the number of processing nodes. As a basis for comparison, *Cura* software is used [4], which is generally used for transforming STL file into G-code. Of course, the processing time decreases in reverse the number of processing nodes. In the table, the case that node is 1 means the result from when *Cura* is used. Thus, we come to know that our devised algorithm shows better processing time.

Node num.	Processing time (*ms*)
1	6000
2	4677
4	2067
8	908
16	428

3 Conclusion

The volume of facet data in a 3D model highly depends on the size or precision of the model and the volume increases to several times. So an efficient distributed algorithm is required. In this paper, we devised a distributed algorithm to convert STL file to G-code for 3D printer. After this work, we will do extensive simulation works and find the other factors to affect performance. And we will also develop our system which can operate on a cloud system. After then, users can get G-code by asking Cloud system to convert their 3D model data.

References

1. 3D printing. https://en.wikipedia.org/wiki/3D_printing
2. Jacobs, P.F.: Rapid Prototyping & Manufacturing: Fundamentals of Stereolithography. Society of Manufacturing Engineers (1992)
3. Pearce, J.M., Blair, C.M., Laciak, K.J., Andrews, R., Nosrat, A., Zelenika-Zovko, I.: 3-D printing of open source appropriate technologies for self-directed sustainable development. J. Sustain. Dev. **3**(4), 17–29 (2010)
4. Cura software. https://ultimaker.com/en/products/cura-software

Cache Aware Web-Based Dynamic Adaptive Streaming Algorithm in Information Centric Networks

Geun-Hyung Kim(✉)

Department of Game and Visual Image Engineering, Dong-Eui University, 176 Eomgwang-no, BusanJin-Gu, Busan 47340, Korea
geunkim@deu.ac.kr

Abstract. Information Centric Networking (ICN) has attracted researchers as a future Internet architecture, because of its content dissemination based on names and ubiquitous in-network caching. Dynamic Adaptive Streaming over HTTP (DASH) has been deployed widely to provide video streaming in the Internet with high Quality of Experience (QoE) reflecting the heterogeneity of network connections and user devices. From reviewing previous studies on the effect of the cache to the dynamic adaptive streaming, pre-fetching of appropriate representations is a key issue. This paper proposes a client assisted dynamic rate adaptation mechanism that could invoke pre-fetching the anticipating representations to in-network cache. Simulation result shows our solution reduces the bit rate oscillations.

Keywords: Information centric networking · Quality of Experience (QoE) · Dynamic adaptive streaming · Content caching · Content distribution · MPEG-DASH

1 Introduction

The advent of new multimedia applications indicates that the contents on the Internet have evolved from textual information to multimedia information [1]. According to [2], global Internet video traffic will be 82 % of all consumer Internet traffics by 2020, up from 70 % in 2015. That is, real-time video applications produce most of the Internet traffic. As users largely care about what content they want opposed to where the content is, the consumption of video content is generously content driven. Current Internet was designed for host-to-host communication where host's IP address is used to present the location of a desired content. The content dissemination on current Internet makes a problem like utilizing content replicas or redundant paths. For efficient content delivery, Information Centric Networking (ICN) has been proposed.

ICN [3] is a novel Internet architecture where communication takes place by exchanging the content, identified by its name, instead of transmitting packets from a content source to a destination. The one of the basic premises of ICN is that adopting the content cache intrinsically and ubiquitously in the network infrastructure would improve performance of content distribution. This in-network caching helps users access the requested content at a nearby position with less request latency.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_167

DASH [4] is an emerging standard for the adaptive streaming that is recently ratified as an ISO/IEC standard to provide good QoE for the video streaming service with considering the heterogeneous nature of network connections, the various network conditions, and the diverse devices. In DASH architecture, an original video content is encoded into multiple representations with various bit rates. Each representation is divided to multiple consecutive media chunks, called segments defined as equally long temporal entities of the response to the DASH client's request. Each segment is assigned a unique URL, an index, and explicit or implicit start time and duration [5]. The relationship between different representations is described in a so-called Media Presentation Description (MPD).

ICN and DASH have several common features though they pursue different goals. The common features are that the content is dealt with in small chunks for the content delivery and the client initiates the content delivery. ICN mainly focuses on efficient, scalable, and secure content distributions through network layer primitives. DASH focuses on QoE guaranteed video streaming by constructing an adaptive set consisting of the segments from representation sets according to the dynamics of service environment. As the increase of volume of mobile video traffic is anticipated, the dynamic adaptive streaming is a native and essential component in the future Internet. In this paper, we propose client assisted dynamic adaptation algorithm for ICN.

The remainder of this paper is organized as follows. We discuss the related works in Sect. 2. In Sect. 3, we describe the proposed system model and performance evaluation. Finally, we have our conclusion and future work in Sect. 4.

2 Related Works

Liu *et al.* [6] introduced a joint client-driven pre-fetching and cache-aware rate adaptation algorithm for DASH. The DASH client cooperates with proxy caches to determine the representation for next requested segments and subsequent segments which the client will most probably request subsequently. However, this proposal required DASH aware caches by implementing the extension protocols. Müller *et al.* [7] proposed an adaptation method involving exponential back-off and network probe in order to estimate the end-to-end path latency and to avoid frequent quality switching caused by proxy cache. Lee *et al.* [8] proposed a video-aware cache server instead of focusing on the client side algorithm that cache server uses traffic shaping to control the content download rate to prevent sudden quality changes. In this algorithm, the cache server should recognize what the representation bit rates users request and shape eligible transfers. Heikkinen *et al.* [9] proposed the CDN architecture, following the upcoming MPEG standard for Server and Network assisted DASH (SAND), to optimize CDN operation for DASH in terms of the caching and the network resource usage as well as the response time. In [9], authors defined the content-aware network element that monitors cache status, exchanges SAND signaling, requests pre-fetching of anticipating segments from the original server, and redirects the client's request to the optimal edge sever dynamically.

Liu *et al.* [10] investigated, using the CCNx test-bed, the performance of caching of Content Centric Networking (CCN), one form of ICNs, in the context of DASH.

Their observation was that the clients can retrieve the video with a higher bit rates than actual bandwidth between the origin server and the client, when the number of client requesting the video is high enough. However, they didn't investigate the DASH rate adaptation logic. Li *et al.* [11] proposed distributed in-network video caching policy for dynamic adaptive steaming to facilitate video delivery over ICN. The proposed video caching policy is a popularity-based decision on what video content should be cached based on minimal coordination of aggregate video statistics dispersively.

Liu *et al.* [12] proposed hop-by-hop adaptive video streaming scheme instead of end-to-end adaptive video streaming of original DASH, to enhance the performance of adaptive video streaming of Scalable Video Coding (SVC) encoded video content in CCN. Wei *et al.* [13] proposed dynamic adaptive streaming solution, employing a self-learning based rate determination strategy and an interest flooding control mechanism, in Content Centric Mobile Network. They defined the load degree of each content carrier and the interest satisfied potential (ISP), reflecting node's capability to handle another node's interest request, which leverages potential field theory and considers load degree and hop count.

Rainer *et al.* [14] derived the upper bounds for performance of Dynamic Adaptive Streaming with idealized caching in Named Data Networking that is taking multi-path transport into account by modeling video streaming as fractional multi-commodity flow problem. They used also SVC encoded video content for their experiments. Their findings show that the buffer-based adaptation logic obtains a higher average video bit rate compared to the rate-based adaptation logic and the rate-based adaptation logic causes the clients to get more bit rates oscillations.

3 Cache Assisted Dynamic Rate Control

In this section, we describe our approach to create a cache assisted dynamic rate adaptation to avoid premature bit rate oscillations.

3.1 System Model

We illustrate the system model considered in our work. We consider a video content is encoded by n different bit rates. So we define a set of representations $\boldsymbol{R} = \{R_1, .., R_i, .., R_n\}$. We denote the average bit rate of representation R_i as r_i, so we define a set of bit rates of representations $\boldsymbol{r} = \{r_1, .., r_i, .., r_n\}$. Each representation consists of m segments. We consider a set of segments of ith representation $S_i = \{S_i^1, .., S_i^j, .., S_i^m\}$ ($i = 1, .., n$). Consequently, a video content for dynamic adaptive streaming consists of $m \times n$ segments. From the content playback perspective, every client requires only m segments among $m \times n$ segments for the content. We also define an adaptation set $\boldsymbol{A}$ to depict a set of segments used for content playback and it forms a sequence of determined segments. Each client can construct its own adaptation set while it playbacks the content according to the environment. In order for each client to construct the best adaptation set, it should make a decision at least m times.

For dynamic rate adaptation, we define available bandwidth, buffer status, and the segment table of neighbors similar to the chunk table in Peer-to-Peer (P2P) system as the environment parameters. When we decide the representation level, we refer these environment parameters. We use an exponential moving average to estimate the available bandwidth $b(k)$ at the representation decision time slot k. In addition, the client measures instantaneous rate b_i to the check the network status, whenever it receives packets. The available bandwidth $b(k)$, with the initial value of $b(k)$ set to zero, is calculated in Eq. (1). We set α to 0.2 to compute the weight average to reduce the impact of temporary fluctuations.

$$b(k) = (1 - \alpha) \cdot b(k-1) + \alpha \cdot b_i \tag{1}$$

The client stores the received segments in the playback buffer, when it receives segments. We denote the playback buffer size in seconds as β and the current buffer level as β_c. After the client starts the playback of video content, playback buffer is drained periodically with a drain rate $r_{out}(\in \boldsymbol{r})$, and the client continuously fills up the buffer by downloading subsequent segment with rate r_{in}. In the QoE perspective, when the buffer is empty or does not have enough data, the playback of video content can be paused and it causes the QoE degradation.

The goal of this paper is to reduce frequent representation switching, that is to increase the correlation of bit rate of segments in the adaptation set, and to maximize the average bit rate of representations for higher video quality. The basic mechanism of representation decision based on the buffer level is depicted in Fig. 1.

When the buffer level decreases to the low level ($<\beta_{low}$), we can infer that the network congestion or cache miss may happen. So, in order to avoid buffer starvation which causes the playback disrupt, adaptation logic decreases representation level

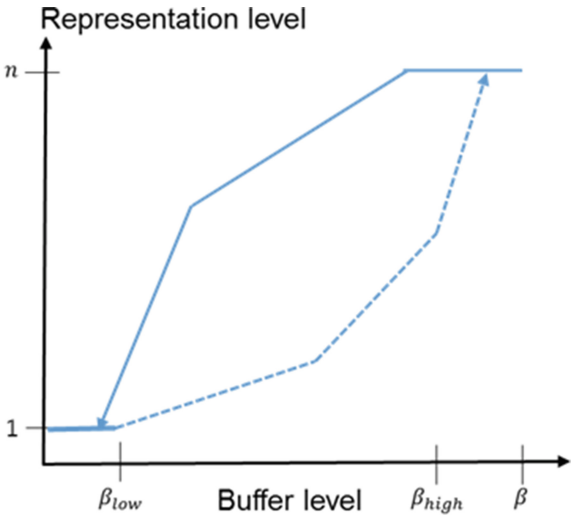


Fig. 1. The conceptual relationship between representation level and buffer level.

Table 1. General dynamics of playback buffer.

Buffer in/out relation	Buffer status
$r_{out} > r_{in}(b_i)$	Buffer level will be decreasing
$r_{out} < r_{in}(b_i)$	Buffer level will be increasing
$r_{out} == r_{in}(b_i)$	Buffer level will be stable

rapidly. When the buffer level increases to the high level ($>\beta_{high}$), we can infer that the available network throughput without congestion may be high or cache hit may happen.

In Table 1, we show the general dynamics of the playback buffer. The buffer drain rate r_{out} and the buffer fill rate r_{in} may be an instantaneous bit rate b_i.

In this paper, we assume that in-network caches are in the network. Therefore, a video segment requests from the clients will be broadcasted to the routers. This interested request includes the representation, which is predicated at the client using DASH technology. Each time before request a segment, DASH would do the rate adaption logic to decide the representation for that segment. The rate adaptation logic periodically monitors the buffer level β_{cur}, available bandwidth $b(k)$, instantaneous bandwidth b_i, and neighbors' segment tables. Basically, the client changes the representation, when the buffer level is change very quickly.

At first, the client refers the neighbors' segment tables to check if which segments are in the in-network caches. If required segment is in neighbors' segment table, the client do not overestimate the available bandwidth. When the client requires the

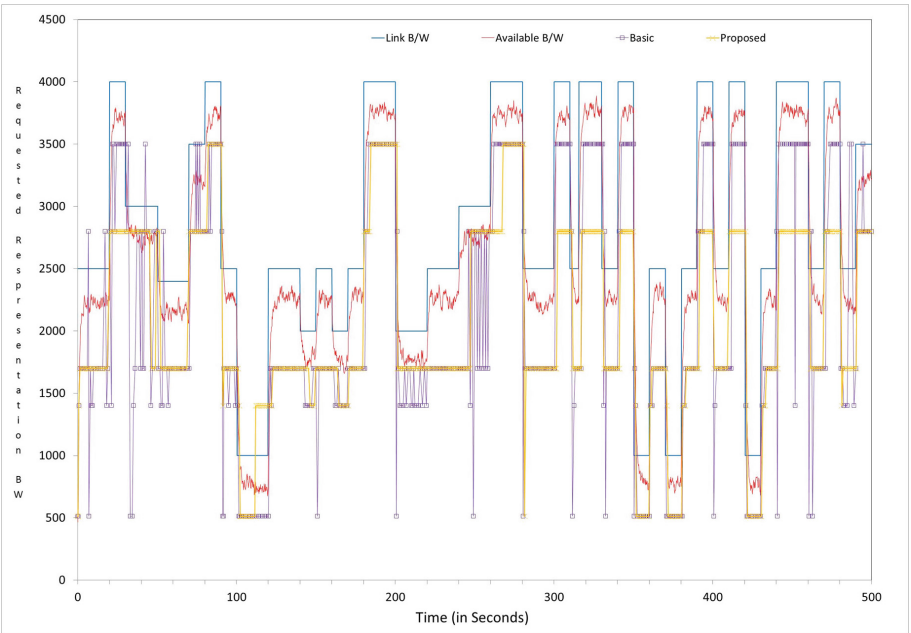


Fig. 2. Representation bitrate for proposed and basic algorithm

segment that the client doesn't know if it is in the cache, the client sends interest messages for subsequent segments in parallel to pre-fetch the segment to the in-network cache and to check if the segment is in the cache.

3.2 Performance Evaluation

In this paper, we perform simple simulation to verify our proposed algorithm using ns-3. For our simulation, we used five representation bitrate: 512 Kbps, 1.4 Mbps, 1.7 Mbps, 2.8 Mbps, 3.5 Mbps. The simulation topology consists of an origin server, a cache and three clients. We set the bandwidth between the origin server and the cache to 2 Mbps and the bandwidth between cache and the client to 5 Mbps. In this experiment, we evaluate the performance proposed algorithm and basic cache under varying bandwidth. The simulation result shows in Fig. 2.

4 Conclusion

This paper aims to reduce the bit rate oscillation and to improve the performance of DASH over ICN. This paper proposes a client assisted dynamic rate adaptation mechanism that could invoke pre-fetching anticipating representations to in-network cache. Simulation result shows our solution reduces the bit rate oscillations.

As the future work, we will derive the optimal solution for increasing the performance of DASH over ICN, carry out QoE subjective experiments, and extend the solution in the large scale simulation.

Acknowledgments. This work was supported by Dong-Eui University Foundation Grant (2014).

References

1. Brito, G.M., Velloso, P.B., Moraes, I.M.: Information-Centric Networks: A New Paradigm for the Internet. Wiley, Hoboken (2013)
2. Cisco White Papers, Cisco Visual Networking Index: Forecast and Methodology, pp. 2015–2020, June 2016
3. Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., Ohlman, B.: A survey of information-centric networking. IEEE Commun. Mag. **50**(7), 26–36 (2012)
4. ISO/IEC 23009-1:2014, Information technology – Dynamic adaptive streaming over HTTP (DASH) Part1: Media presentation description and segment formats (2014)
5. Sodagar, I.: ISO/IEC JTC1/SC29/WG11 W13533 White paper on MPEG-DASH Standard (2012)
6. Liu, C., Bouazizi, I., Hannuksela, M., Gabbouj, M.: Client-driven joint cache management and rate adaptation for dynamic adaptive streaming over HTTP. Int. J. Digital Multimedia Broadcast. (2014)
7. Müller, C., Lederer, S., Timmerer, C.: A proxy effect analysis and fair adaptation algorithm for multiple competing dynamic adaptive streaming over http client. In: Proceedings of VCIP (2012)

8. Lee, D.H., Dovrolis, C., Begen, A.C.: Caching in HTTP adaptive streaming: friend or foe? In: NOSSDAV 2014, pp. 31–36, March 2014
9. Heikkinen, A., Ojanperä, T., Vehkaperä, J.: Dynamic cache optimization for DASH clients in content delivery networks. In: 13th IEEE CCNC, pp. 980–983 (2016)
10. Liu, Y., Geurts, J., Point, J.-P., Lederer, S., Rainer, B., Müller, C., Timmerer, C., Hellwagner, H.: Dynamic adaptive streaming over CCN: a caching and overhead analysis. In: IEEE ICC, pp. 3629–3633 (2013)
11. Li, W., Oteafy, S.M.A., Hassanein, H.S.: StreamCache: popularity-based caching for adaptive streaming over information-centric networks. In: IEEE ICC, pp. 1–6 (2016)
12. Liu, Z., Wei, Y.: Hop-by-hop adaptive video streaming in content centric network. In: IEEE ICC, pp. 1–7 (2016)
13. Wei, Y., Xu, C., Wang, M., Guan, J.: A novel dynamic adaptive video streaming solution in content-centric mobile network. In: 27th IEEE PIMRC (2016)
14. Rainer, B., Posch, D., Hellwagner, H.: Investigating the performance of pull-based dynamic adaptive streaming in NDN. IEEE J. Sel. Areas Commun. **34**, 2130–2140 (2016)

Background Subtraction Framework for Mobile 3D Sensor Data

Seongjo Lee, Seoungjae Cho, Nguyen Trong Hieu, Phuong Chu, and Kyungeun Cho(✉)

Department of Multimedia Engineering, Dongguk University-Seoul, 26 Pildong 3 Ga, Jung-gu, Seoul 100-715, Republic of Korea
cke@dongguk.edu

Abstract. The latest developments in artificial intelligence and sensors have been stimulating research on intelligent service robots and self-driving car technology. Such service robots and self-driving cars must take actions according to the situation by detecting human beings or dynamic environments. This paper proposes a framework that eliminates the background area to facilitate detection of moving robots or vehicles in a dynamic environment. The approach proposed in this paper can eliminate the background at high speed by analyzing an accumulation of 3D data through parallel processing using the voxel data.

Keywords: Background subtraction · Mobile sensor platform · 3D LIDAR

1 Introduction

Along with the recent development of 3D sensors and standardized communication, developments for future industry have been accelerated by the application of these technologies to self-driving cars and cloud robotics to facilitate recognition of the surrounding environment [1]. Methods using 3D sensors can help identify situations in remote areas more accurately and intuitively than existing methods that identify the remote environment using only 2D cameras.

Technologies for detecting changes in a dynamic environment can generate data to enable robots to quickly cope with changes in the environment surrounding disaster areas or with movements of humans or motor vehicles. Such technologies generally use background subtraction. However, it is difficult to eliminate the background because the zero point of a moving sensor varies with the point of view.

This paper proposes a framework that eliminates the background using 3D LIDAR data acquired from moving sensor platforms. To eliminate the background, the proposed approach uses voxel data. The processing time is reduced by reducing the total quantity of 3D data to be processed. Furthermore, the proposed framework design has a structure that can process 3D data in parallel.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_168

2 Related Work

Existing methods for eliminating the background from 3D data generally use sensor data. Shackleton [2] generates cells by segmenting the 3D spatial data acquired by sensors. A specific area consecutively scanned by LIDAR in a relevant cell is determined to be the background. However, 3D LIDAR has low vertical resolution, so it is difficult to scan fixed objects in a remote area in a sequence. Because of this feature, it is difficult to apply it to a broad area.

Kalyan [3] generates a depth map using 3D LIDAR data. Then, the background is separated from the changing whole view in the depth map using a mean-shift filter. However, the mean-shift filter can be applied only to data having zero points that match those of the sensors.

Ortega [4] finds dynamic area from 2D camera images by Gaussian mixture-based background segmentation. Then 3D points from 3D LIDAR are projected to 2D camera image to acquire 3D points corresponding to dynamic area. Ortega minimizes dynamic object detection by using 2D image-based segmentation. However this method works when positions of 2D and 3D sensors are static.

3 LIDAR Simulation System Structure

First, the 3D primitive points acquired by the proposed framework for a frame are built up in the whole area. The whole area is segmented into a grid of spaces of a specific size called "cells." A cell comprises several voxel areas. Figure 1 illustrates the relationships among the whole area, cells, and voxel areas.

Next, the primitive point in each cell processes the scanned voxel data and determines whether the corresponding cell is part of the background or the whole area. As shown in Fig. 1, the algorithm for determining the background can be executed in parallel as each cell is segmented as an independent space. This process helps improve the performance in proportion to the maximum number of cores when using a multicore CPU or a general-purpose graphics processing unit (GPGPU).

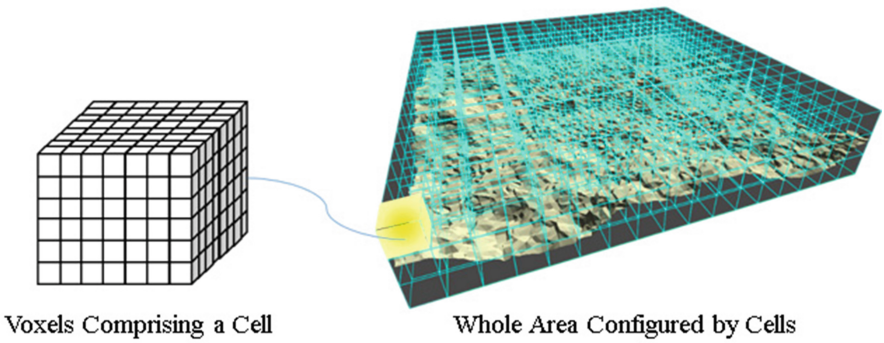


Fig. 1. Hierarchical data structure in the framework.

4 Background Subtraction Approach

For determining whether each cell is part of the background or the whole area, the approach proposed in this paper considers two kinds of features observed in the background in the 3D space that has been built. First, the distribution of primitive points that make up a cell is checked for changes. When an object moves dynamically in a cell, the distribution of the scanned voxels comprising the cell is significantly changed (Fig. 2). To detect such events, the unique value and vector of each voxel in the cell are examined, making use of the change in their distribution caused by the fact that the moving object significantly changes their unique value and vector.

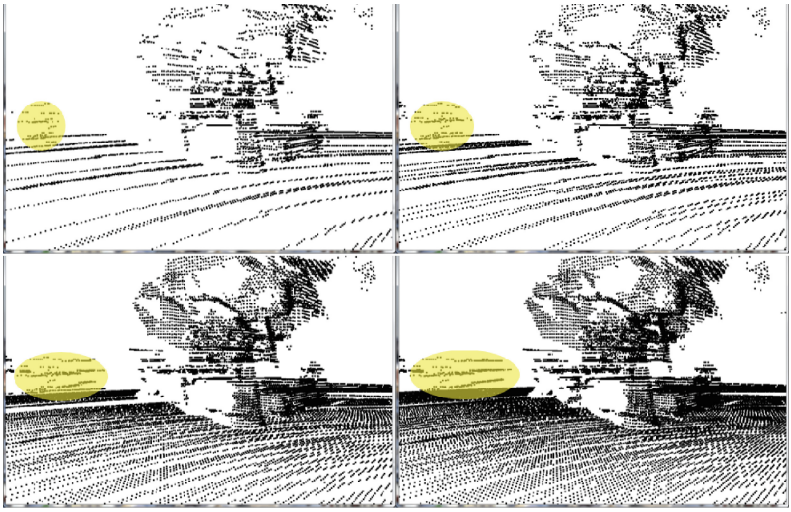


Fig. 2. Change in voxel distribution caused by movement of object.

Next, the change in the height of the scanned voxels in a cell is checked. As shown in Fig. 3, the movement of an object in a cell significantly changes the height of the scanned voxels in the cell. Since voxels are a quantified unit, the difference in the height of the voxels between the previous frame and the present frame in a space of the same size can be quickly estimated.

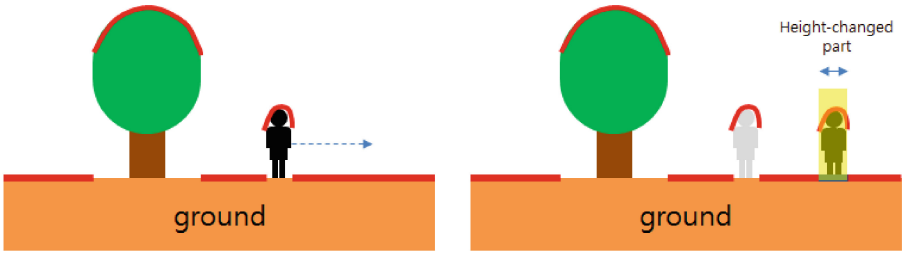


Fig. 3. Change in voxel height caused by movement of object.

5 Implementation and Results

The computer used in this experiment was an Intel i7 4790 processor with 16 GB RAM. The processing rate per frame was 9–23 ms based on a single thread. The voxel visualization program, including the configuration of the proposed framework, was produced using Direct3D. The voxels were visualized by classifying those voxels that were part of background cells, excluding the cells with moving objects by using the relevant program. The dataset used as input to the program was the virtual data generated using the sensor simulator.

Figure 4 presents the implementation results. The picture on the left shows the environment logged on the simulator, and that on the right presents the voxel visualization. The blue area represents the whole area of the corresponding picture, and the black area represents the background. The program did not visualize the voxels above a certain height in identifying the results.

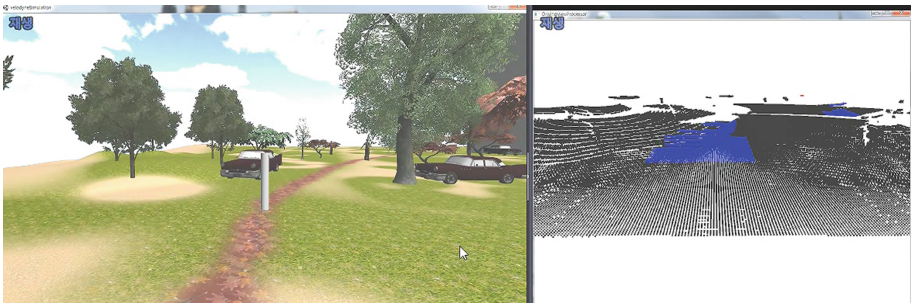


Fig. 4. Implementation results.

6 Conclusion

This paper has proposed an approach for classifying the whole area and the background area of 3D data acquired from moving sensors. The proposed approach builds up 3D points in whole voxel areas, several of which make up a cell, and distinguishes the background from the whole area by processing the voxel data with the primitive points in each cell. To verify the proposed approach, a voxel visualization program was produced, and the whole area and the background area were visualized. The implementation results verified that the whole area and the background area could be classified with sufficient speed.

Acknowledgements. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2015R1A2A2A01003779) and the National Program for Excellence in SW (R7116-16-1014) supervised by the IITP(Institute for Information & Communications Technology Promotion).

References

1. Ko, J.H.: Intelligent robot control for 3D vision technology. Mag. IEEK **38**(12), 72–78 (2011)
2. Shackleton, J., VanVoorst, B., Hesch, J.: Tracking people with a 360-degree lidar. In: Proceeding of the 7th IEEE Conference on Advanced Video and Signal Based Surveillance, pp. 420–426 (2010)
3. Kalyan, B., Lee, K.W., Wijesoma, S., Moratuwage, D., Patrikalakis, N.M.: A random finite set based detection and tracking using 3D LIDAR in dynamic environments. In: Proceeding of the IEEE International Conference on Systems Man and Cybernetics, pp. 2288–2292 (2010)
4. Ortega, A.A., Andrade-Cetto, J.: Segmentation of dynamic objects from laser data. In: Proceedings of the 5th European Conference on Mobile Robots, pp. 115–121 (2011)

A Method for Multi-user Re-identification in Invoked Reality Space

Yunji Jeong, Yulong Xi, Jisun Park, Kyhyun Um, and Kyungeun Cho(✉)

Department of Multimedia Engineering, Dongguk University-Seoul, 26 Pildong 3 Ga, Jung-gu, Seoul 100-715, Republic of Korea
cke@dongguk.edu

Abstract. Natural user interfaces (NUIs) are commonly employed to allow multiple users to interact with each other in a defined space when implementing invoked reality. This has led to the investigation of user identification approaches as the basic technology enabling this interaction. However, the use of a single sensor complicates the accurate extraction of user-specific data because of the limited sensing scope and the occurrence of overlap among users. This paper proposes an approach based on the use of multiple sensors to re-identify users when multiple users repeatedly exit and re-enter the sensing scope.

Keywords: Multi-user Re-identification · Human-computer interface · Invoked reality

1 Introduction

Invoked reality is the technology that enables a user to load virtual content into real-world spaces by using a natural user interface (NUI). Users can employ NUI technology to implement invoked reality space for interaction by applying invoked reality technology to an indoor space. When building the invoked reality space, it is easier to identify a single user and recognize the gesture than to identify and recognize multiple users. However, when multiple users interact in the same invoked reality space, overlapping among users may occur. Then, it becomes difficult to accurately recognize events executed by individual users. Furthermore, the invoked reality space cannot continue maintaining the same indoor environment, because the brightness of indoor lighting may change or the amount of sunlight entering through windows may vary. In consideration of the conditions described above, we applied an approach whereby the features of colors and patterns that are able to expand the sensing scope of sensors are extracted by using multiple sensors in the invoked reality space to minimize the impact of lighting.

This paper describes an approach based on the use of multiple sensors in the invoked reality space to re-identify multiple users. The multi-user identification method could be applied as an invoked reality-based conference system and military drill system based on body gesture recognition in combination with a variety of NUIs.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_169

2 Related Work

NUIs have been developed with the aim of allowing users to transform space into interactive space. Approaches to distinguish among users when multiple users are active in the same space have been actively investigated to ensure that the appropriate service is provided depending on users' locations in the interactive space [1–6]. In this regard, the use of sensors for identifying multiple users is inevitable. In practice, the approach relies on far-infrared sensors, a web cam, and Kinect sensors. Han [1] classified the head, top, and bottom of a user's body by entering their silhouette into the sensing scope using one Kinect sensor. The features of garments were adopted on the assumption that it is highly unlikely that multiple users wearing the same garments would enter the sensing scope. Color and pattern data were extracted using the RGB data of the classified area. A color histogram of extracted data was produced and the difference in the angle between all histograms and the user's final histogram was determined. If the angle difference between both histograms is determined to be smaller than a specified threshold, multiple users are identified.

However, the use of a single Kinect sensor to identify multiple users has limited sensing scope. Thus, an approach involving the use of more than two Kinect sensors to identify users was investigated.

JO [5] configured the interactive space by arranging four Kinect sensors in a space. Users' locations were identified using data of the dimensions of their skeleton. The acquired tracking data and the saved video data were then compared in real time. When the change in the skeleton length was the same, the user was identified as being the same person. Accordingly, multiple users can be identified by mapping each user's ID as 3D data and integrating the data of each user's skeleton.

The following problems were identified in existing research relating to multi-user identification. First, other researchers found it difficult to identify multiple overlapping users [4–6]. Next, it was found to be impossible or difficult to identify users due to changes in illumination [1, 3, 5, 6]. Finally, it was considered difficult to re-identify users who exited and then re-entered the sensing scope of the sensors [4].

In this work we used the following approaches to overcome the above-mentioned problems in existing research. First, the field of view was expanded by using multiple sensors to obtain accurate data even when multiple users obscure each other or overlap with each other. Next, a technology capable of extracting strong color and pattern features was adopted to address changes in illumination to prevent data errors even when the illumination changes. In summary, this paper proposes an approach to identify users who move outside the sensing scope of multiple sensors and re-enter at a later stage as being the same users.

3 User Identification Approach

Our approach to re-identify multiple users is based on the use of multiple Kinect sensors. The re-identification system for multiple users has two kinds of roles: that of server and client. When the user enters the sensing scope, the client extracts the color and pattern features as well as the height of a user according to the position of their

head and the distance between their top and bottom extremities. Several features are extracted because it is difficult to accurately identify users if a single feature is used, particularly when more than two users have very similar features.

The color features of users are extracted such that the impact of changes in illumination is minimized by applying the CIELAB color space. The histogram using the color data of A and B, except for the brightness (L), in CIELAB is used to represent the features. The extraction of pattern features also adopts LBP for minimizing the impact of lighting. LBP indicates '1' when the pixels surrounding the center pixel are brighter and '0' when they are darker than the index value. After applying LBP, the histogram is used as representing the pattern features. In some cases, more than two users may have similarity in terms of both colors and patterns in the invoked reality space. In these cases, trigonometric function features using the distance between a sensor and the user are applied in addition to the user's height. The client transfers all extracted features to the server.

The server integrates the feature data, which it then uses to identify multiple users. The Bhattacharyya distance formula for measuring the similarity in the continuous probability distribution was applied to estimate the similarity of the a and b histograms of CIELAB and the histogram of LBP extracted as above. The closer the estimated value is to '0', the higher the correlation is, whereas an estimated value closer to 1 means lower correlation. A user's height feature is determined by estimating the tangent value by comparing the previous frame with the present frame. The estimation result is used as the user's height feature. The user ID with the highest similarity value and the final similarity value (estimated using (Formula 1)) are transferred to the client. This process enables the re-identification of a user who exits and then re-enters the sensing scope.

$$f(a) = x_1(1 - y_1) + x_2(1 - y_2) + x_3(1 - y_3) + x_4(1 - y_4), (\sum_{n=1}^{n} x_n = 1) \qquad (1)$$

Where n is the number of features, x_1 is the weighed value of CIELAB a, and y_1 is the similarity value of the CIELAB a histogram. Further, x_2 is the weighed value of CIELAB b, y_2 is the similarity value of the CIELAB b histogram, x_3 is the weighed value of LBP, y_3 is the similarity value of the LBP histogram, x_4 is the weighed value of the user's height feature, and y_4 represents the user's height feature value.

4 Experiment

The application was implemented and an experiment was conducted to identify multiple users in the invoked reality space, as proposed in this paper. The experiment involved the use of one server PC and two client PCs. Two 'Kinect for Windows v1' units were connected for extracting and identifying the features of users. The experiment verified that more than two multiple users could be re-identified when re-entering the experimental space. Table 1 presents the identification accuracy that was obtained when multiple users exited and re-entered the invoked reality space.

Table 1. Identification accuracy of users re-entering the space.

	User ID	Client	Identification Accuracy of Client	No. of Tests	Total Identification Accuracy	User Image
Test 1	1	Client1	98%	50	94%	
Test 1	1	Client2	94%	50	94%	
Test 1	4	Client1	74%	50	73%	
Test 1	4	Client2	72%	50	73%	
Test 2	1	Client1	98%	50	97%	
Test 2	1	Client2	96%	50	97%	
Test 2	3	Client1	98%	50	95%	
Test 2	3	Client2	92%	50	95%	

5 Conclusions

This work presented in this paper involved expanding the field of view of sensors by applying multiple sensors to identify multiple users re-entering the sensing scope. Furthermore, feature data was synchronized to enable other sensors to identify users even when one sensor failed to identify a user due to overlapping or users obscuring each other. Moreover, this paper proposed an approach to minimize the impact of changes in the illumination by using the CIELAB color model and local binary patterns as the features. It was verified that multiple users could be re-identified simultaneously by implementing a server-client structure for real-time interaction.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT, and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8501-16-1014) and the National Program for Excellence in SW (R7116-16-1014) supervised by the IITP(Institute for Information & Communications Technology Promotion).

References

1. Han, J., et al.: Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment. IEEE Trans. Consum. Electron. **58**(2), 255–263 (2012)
2. Hu, N., et al.: Multi-user identification and efficient user approaching by fusing robot and ambient sensors. In: 2014 IEEE International Conference on Robotics and Automation (ICRA), pp. 5299–5306. IEEE (2014)
3. Yumak, Z., et al.: Tracking and fusion for multiparty interaction with a virtual character and a social robot. In: SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence, p. 3. ACM (2014)
4. Shimura, K., et al.: Research on person following system based on RGB-D features by autonomous robot with multi-kinect sensor. In: 2014 IEEE/SICE International Symposium on System Integration (SII), pp. 304–309. IEEE (2014)
5. Jo, H., et al.: Motion tracking system for multi-user with multiple kinects. Intl. J. u- e-Serv. Sci. Technol. **8**(7), 99–108 (2015)
6. Satta, R., et al.: Real-time appearance-based person re-identification over multiple kinectTM cameras. In: VISAPP(2), pp. 407–410 (2013)

A Design Scheme of Combined Syllable Fonts for Hunminjeongeum

Jeongyong Byun(✉), Seongbum Hong, and Hoyoung Kim

Department of Computer Engineering, Dongguk University,
123 Dongdae Road, Gyeongju, Gyeongbuk 38066, Republic of Korea
{byunjy, sbhong, hykim}@dongguk.ac.kr

Abstract. Hunminjeongeum can produce about 39.9 billion syllables formed from the combination of relevant fonts are needed in order to represent texts on the computer system. Due to the astronomical number of syllables, the design way is proper for the use of combined-style syllable fonts although fonts are represented in a somewhat less elegant way. In this paper, we analyze the structures of all syllables and propose a font design way that guarantees its completeness by identifying the number of basic glyphs of design target graphemes. And also we will show their efficiency through reduction rate.

Keywords: Hunminjeongeum · Jeongeum · Haerye · Combined syllable font · Hangul jamo · Unicode · KS X 1001 · Grapheme · Design target glyph

1 Introduction

Hunminjeongeum (correct sounds that teach people, called Jeongeum) [1] was created in 1443 and now it is a Korean national treasure and UNESCO Memory of World Register. Hunminjeongeum Haerye(explanations and examples, called Haerye) [1] can produce 39.9 billion syllables [2] by the following fives explanations; the designs of letters, initial sounds, middle sounds, final sounds, and combining letters. The current rules of Korean (Hangul) orthography made in 1933 can generate 11172 modern Hangul syllables. Since Korean Standard Hangul code for information interchange KS X 1001 [3] is able to express no more than 2350 syllables (21 % of modern Hangul syllables), it is a very serious problem. MicroSoft additionally made Unified Hangul code (CP949) to supplement 79 % syllables absent in KS X 1001. Hangul Jamo (alphabet) code in Unicode partially supports about 39.9 billion syllables [3, 4]. KS X 1001 and Hangul Jamo code in Unicode neither follow scientific rules of Jeongeum nor support the full set fonts. Thus we cannot fully express the 39.9 billion syllables in the computer system.

This paper aims at supporting the complete set of fonts that Jeongeum can produce. However, due to the astronomical number of syllables, the design way is proper to use combined-style syllable fonts although elegancy of fonts is a little poor. Then we deduce their formats then produce the number of glyphs as design target and grantee the completeness of combining syllable fonts. We finally show economic benefits and design efficiency of the design method by evaluating reduction rates of each grapheme of syllable characters with both numbers of the graphemes and the glyphs.

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_170

2 Related Work

Many of the previous studies on how to implement Jeongeum on the computer can be divided into three phases. Phase 1 studies [1, 7, 8] is related to scientific research of Jeongeum in which Haerye consists of five explanations and one example. They develop Jeongeum code which is a coding scheme for 17 initial letters, 11 middle letters, and 17 final letters and it becomes the Hangul Jamo code in Unicode with a little modification. The studies related to phase 1 identify that the four rules extracted from Jeongeum would generate 39.9 billion syllables.

The second phase [2, 5] is mainly concerned with investigating some ways to implement the scientific principles of Jeongeum identified in the first phase. For instance, it is intended to develop Jeongeum code system by deciding the object with a relevant coding scheme so as to apply the scientific fundamentals of Jeongeum. This kind of research actually followed the Hangul Jamo code including the properties of Jeongeum that were proposed as a national position to Unicode or ISO 10646BMP [5] in Seoul meeting of ISO/JTC1/SC2 in 1992 and has been finally adopted as one of three kinds of Hangul codes.

The third phase [3, 4, 6] is related to method research of representing all syllables defined in Jeongeum with Hangul Jamo code of Unicode. An optimized font development research [4] is a partial attempt for a combined-style syllable font. This is a desirable idea and an economical strategy. Here the number of glyphs is 1651 using a reduction plan, but it is not clearly proved for full combination of vertical and horizontal middle letters, and the light labial sound characters can be completely represented. It seems that the web input method study [8] is a good practice. The completeness of design target glyph set is not clear in this study.

3 Syllable Structure Analysis

In order to support the full set of fonts for Jeongeum code, we derive basic formats by analyzing syllable structures depending on the shape of middle letters.

3.1 Syllable Structure Formats

Hunmin-Jeongeum created in the 15C is one of phonetic scripts which have an explicit definition of Eumjeol that means a voice unit called syllable. A syllable consists of initial sound(I), middle sound(M), and final sound(F). All Chinese syllables should be composed of the three but the final sound can be empty in Korean.

The initial sound consonants are 17 letters {ㄱ ㅋ ㄴ ㄷ ㅌ ㅁ ㅂ ㅍ ㅅ ㅈ ㅊ ㅇ ㆆ ㅎ ㅿ}. The middle sound vowels are 11 letters {ㆍ ㅡ ㅣ ㅏ ㅓ ㅗ ㅜ ㅑ ㅕ ㅛ ㅠ}. Here the vertical shapes are 5 letters and the horizontal is 6 letters. Since a syllable is composed of vowel at the center and each consonant at the front and the back. The shape of consonant letters can be differently changed by the shapes of the middle vowels and the number of combined characters. Haerye defines the position of the

middle vowel letter that horizontal shape puts below initial letters and vertical shape places at right side of initial letters.

We can classify syllables into 13 formats according to the shape of middle letters to guarantee the least elegancy of syllable fonts. Formats in Fig. 1 have *In, Mn or Fn for n = 1,2,3* which means combining characters, in which all glyphs are spacing character type in vector font. We call *IMv* format 2, *IMvF* format 3, *IMh* format 4, *IMhF* format 5, *IMhMv* format 6, and *IMhMvF* format 7.

Since the middle letters in Jeongeum can be completely composed of two or three characters there are many combinations that modern Hangul doesn't have. For instance, the combination of two middle characters is hh (horizontal-horizontal), hv (horizontal-vertical), vv(vertical-vertical), vh (vertical-horizontal). Here vv belongs to format 2 or 3, hh to format 4 or 5 and hv to format 6 or 7 but vh does not exist in modern Hangul, and hence, we newly call *IMvMh* format 8 and *IMvMhF* format 9. In case of combining three characters, eight combinations occur such that hhh to format 4 or 5, hhv and hvv to format 6 or 7, vhh and format 8 or 9, vh v to format 10 or 11, and lastly hvh to format 12 or 13.

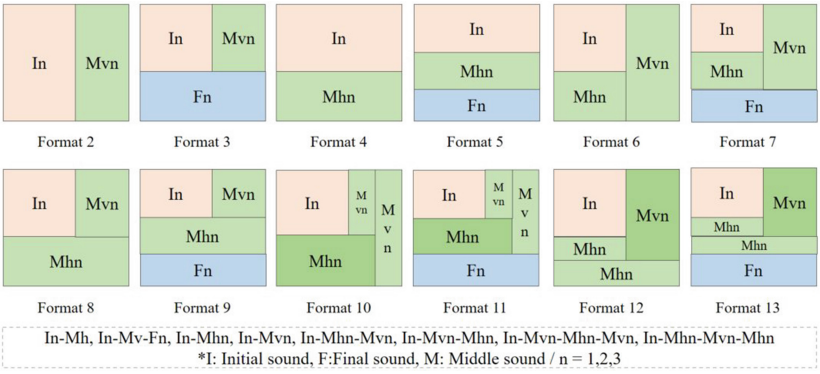


Fig. 1. Formats of syllable structure

3.2 Combination Types of Grapheme

Combination type depends on the initial, middle and final sound letters in combining characters of Hunmin-Jeongeum are able to combine two or three characters. Like the above 13 structural formats, the reason that each font shape is changed should differently design individual character size within the same space. They are classified into three types. Type 1 stands for *Fn, Mvn*, and *Mhn n = 1* and combination type 2 is *n = 2* for them, and Type 3 becomes *n = 3*. The previous studies [6–8] suggest that type 2 and type 3 need to develop one glyph set for each position.

4 Basic Glyphs of Design Target

In this section, we will evaluate the number of design target glyphs based on the three combination types and the 13 formats of syllable structures as you see in Fig. 1.

Format 1. The initial, middle, and final sound letters are spacing characters but do not have syllable structures. The elements of format 1 are extracted from 2 to 13 in Fig. 1 and changed to spacing characters. The initial shapes come from format 2 and the final shapes from format 4. Four mixture types of the middle come from format 6, 8, 10, and 12. The numbers can be gained from the formats. The number of the initials and finals is 204 and that of the middle is 187, and the total comes out to 391 glyphs.

Format 2. The middle sound letters have six horizontal shapes without the finals. The number of middles is 30 glyphs including type 2 and 3. The initials are 102 glyphs. The total comes up to132 glyphs.

Format 3. The middle sound letters have six horizontal shapes together with the finals. The initial shapes are different from those of format 2. There exist 102 final glyphs. The total comes out to 234 glyphs.

Format 4. The middles have five vertical shapes without the finals. The number of middle is 36 glyphs including type two and three. The initials are 102 glyphs. The total comes out to138 glyphs.

Format 5. The middles have six horizontal shapes together with the finals. The initial shapes are different from those of format 3. There exist 102 final glyphs. The total comes to 240 glyphs.

The shape of the initial or final is changed according to the two kinds of mixture combinations of middle sound letters such as horizontal-vertical(hv) of format 6 and format 7 and vertical-horizontal(vh) of format 8 and format 9. The shape of the only middles is also changed but others are not changed from format 8 to format 13.

Format 6. The middle sound letters are hv mixture combination without the finals. Here are two kinds of mixture combinations. One is mixture combination Mv1Mh2 (one horizontal and two vertical vowel) and it gains 27 glyphs. The other stands for Mh2Mv1 and it is 17 glyphs and then subtotal is 44 glyphs. The initials are 102 glyphs. The total comes out to146 glyphs.

Format 7. The middles are hv mixture combination with the final. Here the number of initial and middle glyphs is equal to that of format 6. The size and shape are different from format 6 and it is 44 glyphs. The number of final is 102 glyphs. The total comes up to 248 glyphs.

Format 8. The middle sound letters are vh mixture combination without the final. This kind of combination does not exist in modern Hangul and combination type 3 has vhh and vvh. The initial is equal to that of format 6. The total comes to 44 glyphs.

Format 9. This is format 8 with the final and the initial and final are equal to format 7. The total comes out to 44 glyphs.

Format 10. The middle sound letters are vhv mixture combination without the final. This does not exist in modern Hangul and the number of glyphs is 16. The initial is equal to format 6. The total comes out to 16 glyphs.

Format 11. This is format 10 with the final and the initial and final are equal to format 7. The total comes to 16 glyphs.

Format 12. The middle sound letters are hvh mixture combination without the final. The number of the middle is 17. The initial is equal to format 6. The total comes out to 17 glyphs.

Format 13. This is Format 12 with the final and the initial and final are equal to format 7. The total comes up to17 glyphs.

5 Mixture Combination of Middle Letters

In this research, we extracted 13 formats and 3 combination types. The research focuses on vh or hv mixture combination of the middle with or without the final in the syllable, because the shape of middle letters makes that of the initial or the final change. Format 1 shows every case of combining the middles. The same kind middles come from format 2 or 4 different kinds comes from the others. Even numbers of format stand for syllables without the final and odd numbers are syllables with the final.

There are two types of mixture combination. One is a hv type which includes hvv, and hhv, hvh in both modern Hangul and Jeongeum. The other is a vh type including vh, vhh, vvh, and vhv which only belonged to Jeongeum. Examples of vh type are presented in Fig. 2-(a)(b)(c) for format 8, Fig. 2-(d)(e)(f) for format 9, vhv type as Fig. 2-(g)(h) for format 10–11, and hvh type in Fig. 2-(i)(j) for format 12–13.

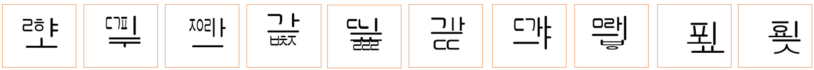


Fig. 2. Examples of mixture middles: (a–f) vh-types, (g, h) hvh & (i, j) vhv

6 Evaluation

In this research, we evaluate the number of graphemes depending on Jeongeum principle of Haerye. Σ*In for n* = 1,2,3 means combining two or three initial consonants and then it results in 5219 characters. The number of the middle from Σ*Mn* is 1413 characters and Σ*Fn* becomes 5220 characters including the empty final consonant cases and the total of three graphemes come out to 11,902. The total syllables H from Eq. (1) result in 39,856,772,340, that is, about 39.9 billion.

$$H = \sum In * \sum Mn * \sum Fn \, for \, n = 1, 2, 3 \tag{1}$$

Table 1. Reduction rate

Graphemes	Haerye	Glyphs	Rate (%)
Initial	5,219	714	86.3
Middle	1,463	561	61.7
Final	5,220	408	92.2
Total	11,902	1,683	85.9

The total of design target glyphs, which are shown as in Table 1, comes out to 1683 glyphs evaluated from Sect. 4. In the below Eq. in (2), g stands for the number of design target glyphs and h means that of graphemes from Eq. (1). While evaluating the expression to get reduction rate, we can confirm that the reduction rate for the initial, the middle, and the final turn out to be 86 %, 62 %, and 92 %, respectively. Total reduction rate is 85.9 %.

$$R = 100 - (g/h)*100 \tag{2}$$

7 Conclusions

In the present study, we propose a way of developing economical fonts which can represent about 39.9 billion syllables depending on Haerye. As the number of Jeongeum syllables is astronomical, we selected the method of designing a combined-style syllable fonts which is not somewhat elegant. We identified 3 combining types and 13 formats from syllable structures defined in Haerye and evaluated the number of design target glyphs to develop the combined-style syllable fonts. We also showed that the proposed method guarantees the representation of all syllables of Jeongeum.

In comparing the number of graphemes of Haerye to that of design target glyphs for combined syllable fonts, we can conclude that each reduction rate has been improved significantly, and specifically, that of final consonant set reaches up to 92 % and the total reduction rate comes to 89.5 %. This indicates that the proposed method is greatly effective.

In the future research, we will closely examine Jeongeum characteristics and requirements of designing elegant fonts to reduce the number of design target glyphs as well as some related elements. Based on this, we will make an attempt to achieve a maximum efficiency at a minimum.

Acknowledgments. This work was supported by the National Research Foundation in Korea (NRF-2015R1D1A1A01060408).

References

1. Kang, S.: Hunminjeongeum Research, Sungkyunkwan University publishing division (1991)
2. Byun, J.: An improvement of hangul code for implementing Hunminjeongeum principle. KIISE Rev. **12**(8), 72–76 (1994)
3. KS X 1026–1:2007 A Guide on Hangul Processing for Information Interchange, Korea Agency for Technology and Standards (2007)
4. Byun, J., Kim, K.: An optimized font development of graphemes by Hunminjeongeum principle. In: Proceedings of 2007 KIPS Spring Conference, pp. 690–693 (2007)
5. Codechart, Unicode (2016). www.unicode.org
6. Byun, J., Oh, S.: A representation of Korean syllable characters defined in hunminjeongeum. In: Cham, T.-J., Cai, J., Dorai, C., Rajan, D., Chua, T.-S., Chia, L.-T. (eds.) MMM 2007. LNCS, vol. 4352, pp. 662–670. Springer, Heidelberg (2006). doi:10.1007/978-3-540-69429-8_72
7. Byun, J.Y., Lim, H.C.: A legislation principle of Hunminjeongeum creation and hangul code system. In: Proceedings of The 3rd Domestic Conference on Hangul and Korean Language Processing, KIISE, pp. 155–158 (1991)
8. Byun, J.Y., Lee, H.J.: A web input method for full set syllables defined by Hunmin-jeongeum. J. KIISE Comput. Pract. Lett. **19**(6), 371–375 (2013)

Proposal of a Resource-Monitoring Improvement System Using Amazon Web Service API

Kyu Ik Kim[1], Musa Ibrahim M. Ishag[2], Myungsic Kim[1], Jin Suk Kim[1], and Keun Ho Ryu[2(✉)]

[1] NeoForce Co., Ltd.,, Seoul, South Korea
han_bando@naver.com,
ora011@nate.com, 0210kimjs@daum.net
[2] School of Electrical and Computer Engineering,
Chungbuk National University, Cheongju, South Korea
{ibrahim, khryu}@dblab.chungbuk.ac.kr

Abstract. Cloud computing uses virtualization technology to segment mass computing resources into various service types, such as Infrastructure as a Service, Platform as a Service, and Software as a Service. The system manager does not need to manage hardware, resource expansion and reduction are convenient, and management is not restricted by time or space. The initial cost for installing mass servers and establishing service infrastructure can be effectively reduced. Amazon, IBM, and Microsoft are the main global vendors that provide the cloud computing service. They globally distribute and manage high-availability cloud service resources. However, when resources generated by users are in different management regions, they cannot be integrated and monitored in the management page provided by the service vendor. In this study, a resource-monitoring service using an API provided by the Amazon Web Service (AWS) was implemented; however, the reception latency time occurred continuously because information is received through the API every time the page changes from the actual service step. To solve these problems, the current condition collector that collects resource information was linked with database management system to enhance the performance by reducing unnecessary requests of calling the AWS API.

Keywords: Cloud computing · System monitoring

1 Introduction

Cloud computing uses virtualization technology to segment mass computing resources into various service types, such as Infrastructure as a Service Infrastructure as a Service, Platform as a Service, and Software as a Service. Thus, the user can use resources such as server, storage, and software anywhere and anytime. Server managers can reduce the initial cost for establishing servers by using cloud computing services; maintenance management due to failure and aging of server equipment has become unnecessary. In addition, server managers can flexibly expand/reduce resources

© Springer Nature Singapore Pte Ltd. 2017
J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*, Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9_171

necessary for service operation so that the rapid increase in resource usage in each server can be flexibly managed [1, 2].

Moreover, the prestigious US information technology research and consulting company Gartner announces 10 strategic technologies to lead the future every year. Figure 1 shows the 10 strategic technologies announced by Gartner in 2015, in which cloud computing, which started in 2009, was selected as a strategic technology and positioned as a core IT technology [3].

Global firms such as Amazon, IBM, and Microsoft provide cloud computing resources that can be used globally, and services to manage these resources are operated by each vendor. Users can use various solutions provided by vendors by using management services, and resources of each solution can be added/eliminated to check the status and properties. Moreover, each vendor has distributed data centers in places worldwide for offering cloud computing services, and users can generate resources according to each distributed region.

However, when the resources generated by users are in different management regions, they cannot be integrated and monitored in the management page provided by the service vendor. If resources not used by the service manager using the cloud computing service are not removed, unnecessary costs are charged in which a list of all used resources must be quickly grasped. Therefore, we have implemented a monitoring system that can check the current condition and status of resources used in the cloud computing environment. However, the created system uses an API in which a cloud computing service vendor calls to acquire information every time a request occurs. This increases the latency time for reception of all requested results, and the API has the feature of requesting resource information output again within a short time.

To improve the problems that occur in the existing system that directly uses API provided by the Amazon Web Service (AWS), a structure that interconnects a database management system (DBMS) and a current condition collector, which collects resource information using this API, are proposed in this paper. The performance of the proposed system is improved by shortening the reception time for resource-list requests in the monitoring system.

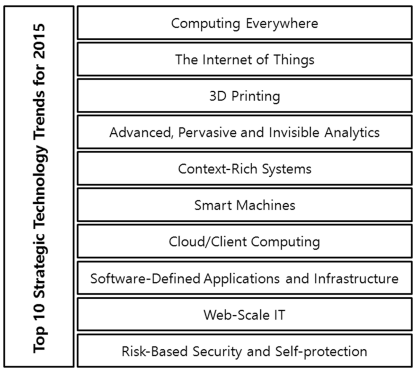


Fig. 1. Top 10 Strategic Technology Trends for 2015.

2 Related Work

2.1 Amazon Web Service

The AWS started in 2006, presenting cloud computing as a web service form to provide IT infrastructures to firms. The current AWS provides expandable and stable infrastructure cloud platforms and its data centers are located in USA, Europe, Brazil, Singapore, Republic of Korea, Japan, and Australia. Amazon Elastic Compute Cloud (EC2), Amazon Relational Database Service (RDS), and Amazon Simple Storage Service are representative services. Various services are provided to establish data centers and new services are continuously added [4].

2.2 Datadog

Datadog is a SaaS-based software and IT system monitoring solution that provides performance monitoring and analysis tools on products such as Docker, MySQL, and AWS [5]. A separate AWS API is not used to call information from Web Application Server (WAS) because information related to monitoring the Datadog server is collected by installing the agent in the device in which the software is installed. When the user logs into the management console, the collected data is analyzed and information is provided.

2.3 New Relic

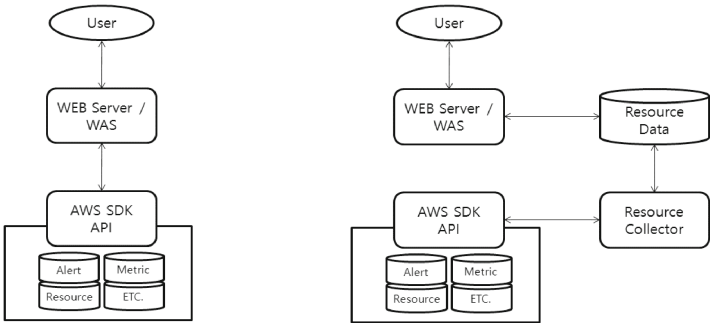
A SaaS-based monitoring solution systematically manages software and monitors various products such as applications, synthetic browsers, servers, and mobiles [6].

3 System Design and Experimental Results

3.1 System Design

We implemented a structure to obtain resource information from the system before AWS SDK API is directly called from WAS. However, loads attempting direct calls of AWS SDK API from the WAS server increased as AWS users increased. This directly influenced service quality. Figure 2(a) shows the structure of the existing system.

To increase service quality by reducing WAS load and providing most of the current data in the monitoring system, we implemented a collector that periodically collects and saves AWS resource information. The recency of resource information can be maintained when the collection period of the collector is shorter. However, change of meta-related information in resources in a short period is rare; thus, the information collection time of the collector was defined as 2 h in this study. Figure 2(b) show the structure of the proposed system. In the existing structure, the API and WAS directly communicate, and WAS uses collected information. In contrast, the proposed structure calls the API and collects information.



(a) Previous Structure (b) Proposed Structure

Fig. 2. System designs.

3.2 Experimental Environment

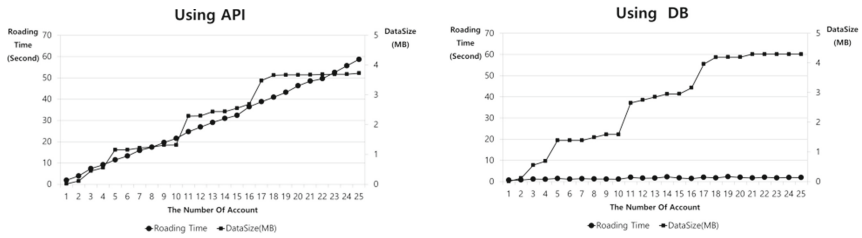
Regarding the experimental environment, EC2 and RDS services provided by AWS were used in WAS and DBMS, and the Chrome browser was used with a Window environment for the client to connect to the server. Table 1 shows the environmental specifications of the experiment.

Table 1. Experimental environment

Specifications		
AWS/EC2	Instance Type	M4.large
	OS	Ubuntu 14.01
	JAVA	Java 1.7.0_91
	WEB/WAS	Tomcat 7.0.52
AWS/RDS	Instance Type	M3.large
	DBMS	MySQL 5.6.27
Test Device	CPU	Intel i7 @2.4 Ghz
	Memory	Samsung 8 GB
	Storage	Samsung SSD 128 GB
	OS	Windows 10 x64
	Browser	Chrome (Ver. 51.0)

3.3 Experimental Results

We measured the time used in the current information page with the most resource information calls in the earlier implemented monitoring system to compare the improvement level. Twenty-five AWS accounts were used in the experiment, and the information was acquired for each account. The experiment was repeated five times with each method to measure the time used and calculate the mean value. Figure 3 shows the change in the time used in the results of the information-acquisition experiment.



(a) Result of Previous Structure (b) Result of Proposed Structure

Fig. 3. Experimental results

Figure 3 shows that for acquiring information from all 25 accounts, the existing method in which AWS API is directly called used 58.69 s and the proposed system structure uses 1.87 s, with approximately 57 s improvement.

4 Conclusion

With the increase in cloud computing and usage of cloud services, the necessity of monitoring systems for cloud resources has increased. We used AWS API of a large cloud computing service vendor to implement a monitoring system that can integrate and show information of accounts and managed regions. However, we identified a limitation in gaining resource information of many accounts in the Web-based management page through the API, and proposed a collector to collect the current condition of resources and improve the system. The experimental results show that the performance was improved by approximately 57 s compared to the existing method in the equivalent condition by using 25 AWS accounts. The time of information acquisition was maintained even when the number of accounts changed during the inquiry.

In the proposed information collector of the current condition of resources, information is collected every 2 h, and information gaps can occur. Therefore, in our future research, we will attempt to solve this problem by using a method that can maintain the recency of information by separately updating resources by interconnecting the event log that is provided by AWS.

References

1. Lee, B.Y., Park, J., Yoo, J.: Implementation of data processing of the high availability for software architecture of the cloud computing. J. Korea Contents Assoc. **3**(2), 32–43 (2013)
2. Armbrust, M., et al.: A view of cloud computing. Commun. ACM **53**(4), 50–58 (2010)
3. Gartner: Gartner identifies the top 10 strategic technologies for 2015 (2016). http://www.gartner.com/
4. Amazon Web Service (2016). https://aws.amazon.com/
5. Datadog (2016). https://www.datadoghq.com/
6. New Relic (2016). https://newrelic.com/

Author Index

A
Abebe, Mesfin, 680
Ahmed, Ejaz, 694
Ahn, Byeong-Eon, 932
Ahn, Hyejung, 16, 600
Ahn, Sang-Il, 579
Alfian, Ganjar, 16, 600
Altaha, Mustafa, 900
An, Euiri, 717

B
Back, Moon-Ki, 619
Back, Seungchan, 674
Bae, Changseok, 779
Bae, Han-Chul, 74, 97
Bae, Ji-Hye, 786, 793
Baek, Du-San, 950, 957
Baek, Seongmin, 29
Bai, Hongtao, 449, 471, 477
Bang, Green, 360, 555
Bong, Wonwoo, 963
Byun, HwiRim, 920
Byun, Jeongyong, 1096

C
Cao, Jiayi, 352, 456
Cao, Yiquan, 352, 456
Cao, Yuling, 236, 421
Cha, Byung Rae, 937
Cha, Guang-Ho, 104
Cha, Jea-Hui, 23
Cha, Young-Tae, 74
Chao, Han-Chieh, 81
Charoenseang, Siam, 3
Chen, Fangpeng, 332
Cheng, Jingde, 87
Cheon, Jae-Yoon, 531
Cho, Buseung, 799
Cho, Dae-Soo, 932
Cho, Hsin-Hung, 81
Cho, Kyungeun, 61, 1086, 1091
Cho, Seoungjae, 1086
Cho, Soohyun, 976
Cho, Wonhee, 389
Cho, Yeongpil, 694
Cho, Young-Hwa, 1067
Choe, Jong-Won, 607
Choi, Dong Hag, 699
Choi, Eun Mi, 637
Choi, Eunmi, 389
Choi, Haechul, 944
Choi, Hoon, 230
Choi, Hyorin, 674
Choi, Jae Sung, 660, 712
Choi, Jae-Young, 1067
Choi, Jang-won, 1031
Choi, Ji-Hyen, 607
Choi, Ji-Soo, 517
Choi, Keun-Chang, 224
Choi, Mijin, 867
Choi, Moon Jong, 712
Choi, Yoo-Joo, 665
Choi, YunSeok, 806
Choo, Hyun-lock, 97
Chu, Phuong, 1086
Chung, Jaehwa, 211
Chung, Man-Soo, 705
Chung, Mokdong, 926
Chung, Myoungbeom, 9
Chung, Tae-Sun, 861, 867
Chung, Tai-Myoung, 579, 594, 1019
Chung, Yuk Ying, 779

D
Doo, Won-Sik, 579
Dou, Zheng, 253, 259, 266

© Springer Nature Singapore Pte Ltd. 2017

J.J. (Jong Hyuk) Park et al. (eds.), *Advances in Computer Science and Ubiquitous Computing*,
Lecture Notes in Electrical Engineering 421, DOI 10.1007/978-981-10-3023-9

F
Feng, Qingsong, 259
Feng, Xin, 421

G
Gao, Hongbiao, 87
Gao, Yan, 463
Gil, Joon-Min, 273, 490, 1031
Gil, Myeong-Seon, 837
Guo, Lili, 253

H
Hague-Chung, 224
Han, Chang Ho, 963
Han, Kihong, 637
Han, Seok-Hyeon, 160
Han, Youn-Hee, 831, 849
Hao, Fei, 730
He, Lili, 449, 463, 471, 477
Heo, Joo-Seong, 831, 849
Heo, Yoon-A, 185
Hieu, Nguyen Trong, 1086
Ho, Jonggab, 997, 1013
Hong, Jang-Eui, 654, 912
Hong, Sangjin, 686
Hong, Seongbum, 1096
Hong, Sung-Hwa, 490
Hu, Chengquan, 463
Hu, Hanfeng, 332
Huang, Kaisi, 61
Huang, Shih-Yun, 81
Huh, Jun-Ho, 1046
Hwang, Humor, 900
Hwang, Hyun Cheon, 982
Hwang, SangWon, 736

I
Igorevich, Rustam Rakhimov, 894
Iida, Hiroyuki, 66
Ishag, Musa Ibrahim M., 1103

J
Jang, Gyujin, 430
Jang, Jong-Wook, 23, 143
Jeon, Seunghyub, 436
Jeon, Yong-Tae, 793
Jeon, You-Boo, 699, 742
Jeong, Seokhoon, 988
Jeong, Wooseong, 132
Jeong, Young-Sik, 160, 185, 371, 920
Jeong, Yunji, 1091
Ji, Huan, 352, 456
Ji, Sai, 312, 323, 332, 342
Jiang, Shanshan, 477
Jiang, Yu, 449, 463, 477
Jin, Rize, 867
Jo, In-Jae, 665
Joe, Dong-Wan, 279
Joe, Inwhee, 717
Jong-Hwan, Beck, 768
Ju, Yong-Wan, 279
Jun, Moon-Seog, 205, 215, 224
Jung, Doo-Hee, 762
Jung, Euihyun, 173, 649
Jung, Hwayung, 997
Jung, Jun-Kwon, 594
Jung, Sungin, 436
Jung, Woosung, 654, 912
Jung, Yong-hwan, 1031

K
Kang, Jungho, 221
Kang, Ju-Sung, 567
Kang, Kyoungwoo, 42
Kang, Myung Ju, 906
Kang, Soo Hyeong, 812
Kang, YunHee, 42, 906
Kim, Bongjae, 573, 723
Kim, Bo-Ram, 705
Kim, Bumryoung, 215
Kim, Dae-Young, 149, 371
Kim, Donggyu, 117, 883
Kim, Dong-Hee, 579, 1019
Kim, Dongkyun, 799
Kim, Doohwan, 654
Kim, Eun, 484
Kim, Eung Tae, 549
Kim, Geun-Hyung, 1079
Kim, Hoyoung, 1096
Kim, Hwan-Kuk, 74, 97
Kim, Hyungheon, 855
Kim, Hyung-Joo, 221, 888
Kim, Hyun-Woo, 160
Kim, In-Hwan, 793
Kim, Jaesoo, 221
Kim, Jeong Gon, 812
Kim, Jeong-Dong, 573, 723
Kim, Jeong-Ho, 205
Kim, Jihun, 944
Kim, Jin Su, 937
Kim, Jin Suk, 1103
Kim, Junsu, 762
Kim, Kangho, 436
Kim, Kisub, 912
Kim, KyeYoung, 932
Kim, Kyong-Han, 849
Kim, Kyoung-Han, 831
Kim, Kyu Ik, 1103

Kim, Kyung Yeul, 750
Kim, Mansik, 221
Kim, Mi Ryang, 407
Kim, Mi, 497, 1039
Kim, Min-Gu, 937
Kim, Mi-Sun, 360
Kim, Moon-Hyun, 198, 401, 430
Kim, Myung In, 756
Kim, Myunggyu, 29
Kim, Myungsic, 1103
Kim, Noo-ri, 806
Kim, Pyeongkang, 855
Kim, Pyung Soo, 549, 812
Kim, Sanghoon, 246
Kim, Sangyoon, 198
Kim, Seokhoon, 149, 366, 371, 484
Kim, Soon-Kyeom, 654
Kim, Su Min, 762
Kim, Sungsoo, 861
Kim, Sungsuk, 1074
Kim, Sungyun, 484
Kim, Sunho, 971
Kim, Suntae, 680
Kim, Tae-Hyoung, 23
Kim, Taewoo, 855
Kim, WooSik, 906
Kim, Yong, 750, 756
Kim, Yonghoon, 926
Kim, Yong-hwan, 799
Kim, Yoon Ho, 376
Kim, Young-Jong, 517
Kim, Young-Nam, 401
Ko, Il-Ju, 9, 360, 555, 561
Ko, Jong-Won, 1067
Ko, Kuk Won, 414, 963, 988
Ko, Kwangman, 694
Ko, Seok-heon, 819
Ko, Soong-ho, 42
Koh, Kwangwon, 436
Kong, Ki-Sik, 1053
Koo, Dong-Su, 631
Ku, Tai-Yeon, 230
Kwon, YongJin, 166

L

Lee, Byeong Rae, 982
Lee, Byungjeong, 674, 950, 957
Lee, Dongseok, 861
Lee, Eung Hyuk, 549
Lee, Gangin, 873
Lee, Haeng-gon, 1031
Lee, Hyun, 660, 712, 786, 793
Lee, Hyung-ju, 1031
Lee, HyunSoo, 806
Lee, Jae Seung, 211
Lee, Jaeho, 16, 111, 137, 511, 600
Lee, Jaekwon, 912
Lee, Jaewon, 717
Lee, Jee-Hyong, 806
Lee, Jeho, 762
Lee, Jeong Hyu, 680
Lee, Jeunwoo, 779
Lee, Ji Yeon, 414, 963, 988
Lee, Joohun, 665
Lee, Joong-Yeon, 525, 531
Lee, Judae, 878
Lee, Jun-dong, 819
Lee, Jung-Min, 712
Lee, Jung-Won, 674, 950, 957
Lee, Kangwoo, 971
Lee, Keun-Ho, 742
Lee, Ki-Hyun, 625
Lee, Kwangil, 888
Lee, Kyu-Chul, 619
Lee, Kyungrak, 717
Lee, Nam Young, 1025
Lee, Nam-Yong, 497, 505, 517, 525, 531, 1005, 1039
Lee, Samuel Sangkon, 442
Lee, Sangjoon, 414, 963, 988, 997, 1013
Lee, Seongjo, 1086
Lee, SeoukJoo, 1059
Lee, Seungwon, 944
Lee, Soo-Hwan, 579, 594
Lee, Yonghee, 971
Lee, YongSeok, 736
Lee, Yun-Seok, 484
Leelathakul, Nutthanon, 154
Li, Chao, 253
Li, Chunmei, 253, 259
Li, Hua, 236, 421
Li, Mengxuan, 61
Li, Tong, 298
Li, Tonglin, 449
Li, Yan, 36
Lim, Hyun-Kyo, 831, 849
Lim, JongBeom, 273
Lim, Sora, 166
Lim, Suk-Young, 932
Liu, Dengzhi, 323, 342
Liu, Jing, 54, 285, 291, 306
Liu, Qi, 312, 332
Liu, Xiaodong, 312

M

Ma, Xingkong, 54
Min, Dugki, 637, 843, 894
Min, Se Dong, 699, 997, 1013

Miranda Lopez, Erik, 376
Moon, Daejin, 932
Moon, Nammee, 686
Moon, Seo Yeon, 396
Moon, Yang-Sae, 837
Myeong-Suk, Pak, 768

N
Na, Yeji, 997, 1013
Nam, YoungKwang, 736
Namgung, Jung-Il, 774
Nguyen, Tuan Anh, 637
Nkenyereye, Lionel, 143
Nsaif, Saad Allawi, 585

O
Oh, Gil-Tak, 619
Oh, In-Taek, 525, 531
Oh, Jung-Hoon, 619

P
Paek, Yunheung, 694
Pan, Sung Bum, 937
Panjan, Sarut, 3
Panumate, Chetprayoon, 66
Park, Boo-Kwang, 920
Park, Byeong-Seok, 185
Park, Doo-Soon, 366, 699, 730
Park, Geunil, 215
Park, Hojoong, 567
Park, Jae-Hyeon, 950
Park, Jeong-Ho, 723
Park, Ji Su, 982
Park, Jinhee, 430
Park, Jin-Ho, 497, 505, 517, 525, 531, 1005, 1025, 1039
Park, JinSoo, 699
Park, Jisun, 1091
Park, Jiyoung, 613
Park, Jong Hyuk, 376, 380, 385, 396
Park, Joon Ho, 505, 1025
Park, Sang-Hyeon, 205
Park, SangIl, 1059
Park, Seok-Cheon, 279
Park, Seong-Min, 74
Park, Soo-Hyun, 774
Park, Young B., 625, 631
Punyawee, Anunpattana, 66

R
Rhee, Jong Myung, 585
Rhee, Jongtae, 16, 600
Rimcharoen, Sunisa, 154
Rui, Xuhua, 843
Ryang, Heungmo, 122, 127
Ryu, Keun Ho, 1103

S
Sakamoto, Kei, 48
Sanggug, Park, 543
Sang-Hoon, Kim, 768
Sasaki, Mayumi, 48
Seo, Bang Won, 812
Seo, DongBum, 742
Seto, Yoichi, 48
Sharma, Pradip Kumar, 380, 385
Shen, Jian, 323, 342
Shih, Timothy K., 81
Shim, Jae-Sung, 279
Shin, Byeong-Seok, 36
Shin, Hwa Seon, 613
Shin, Jong-Hwan, 957
Shin, Jung-Hoon, 680
Shin, Sanggyu, 48
Shin, Seul-Ah, 517, 1005
Shin, Soo-Young, 774
Shin, Yoonmo, 16
Shon, Jin Gon, 982
Si, Guangzhen, 253, 259
Sim, Dae-Soo, 730
Singh, Saurabh, 380, 385, 396
Son, Bong-Ki, 111, 137
Son, Siwoon, 837
Song, Liang, 61
Song, Myeong-Uk, 594
Song, Wei, 61
Srikamdee, Supawadee, 154
Suk, Sangkee, 211
Sung, Bokyung, 561

T
Truong, Mai Thanh Nhat, 246

U
Uddin, Md. Zia, 407
Um, Kyhyun, 1091

W
Wang, Changwon, 997, 1013
Wang, Jin, 285, 291, 298, 306, 352, 449, 456, 477
Wang, Jinyu, 266
Wang, Kai, 463
Wang, Xue, 471
Wang, Yongjun, 54
Wang, Yu-Zen, 81
Wang, Zhen, 312
Wang, Zhilong, 179

Won, Hee-Sun, 230
Woo, Sang-Uk, 1019
Wu, Hao, 332

X
Xi, Yulong, 61, 1091
Xiao, Zi, 266
Xie, Peidai, 54
Xu, Chao, 236

Y
Yang, Gil-mo, 819
Yang, Huamin, 236
Yang, Seokwoo, 837
Yang, Sun Ok, 1074
Yeom, Sun-Ho, 774
Yeom, Yongjin, 567, 825
Yoo, Cheol Jung, 680
Yoo, Dongchang, 825
Yoo, Hee-Kyung, 680
Yoon, Chui Young, 192
Yoon, Soojin, 97
Yoon, Yong-Ik, 607, 705
Youn, Jonghee M., 944
Yu, HeonChang, 273
Yu, Hye-Yeon, 401
Yu, Xiaofeng, 352, 456
Yun, Unil, 117, 122, 127, 132, 873, 878, 883

Z
Zhang, J.W., 285, 291, 306
Zhang, Tao, 179
Zhao, Yang, 266
Zheng, Y.H., 285, 291, 306
Zheng, Yuhui, 298
Zhou, Xiaozhou, 298
Zhou, Yuan, 87